



INDIAN AGRICULTURAL
RESEARCH INSTITUTE NEW DELHI

18875

L A R I S

MGIC-84-10 AR-21 6 40-1 000

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A—MATHEMATICAL AND PHYSICAL SCIENCES

VOL CLX

LONDON

**Printed and published for the Royal Society
By the Cambridge University Press
Fetter Lane E.C.4**

15 June 1937

18875

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOL CLX

No. A 900—1 May 1937

	PAGE
The Inverse Law of Gravitation—II By E A Milne, F R S	1
The Inverse Square Law of Gravitation—III By E. A. Milne, F.R S	24
A Theoretical Formula for the Solubility of Hydrogen in Metals By R. H. Fowler, F R S and C J Smithells	37
Complex Variables in Quantum Mechanics By P. A. M. Dirac, F.R S	48
On the Pattern of Proteins By D. M. Wrinch. Communicated by R. Robinson, F.R S.	59
The Relationship between Viscosity, Elasticity and Plastic Strength of a Soft Material as Illustrated by some Mechanical Properties of Flour Dough. IV—The Separate Contributions of Gluten and Starch. By R K Schofield and G W. Scott Blair Communicated by Sir John Russell, F R S.	87
The Quantum Theory of Atomic Polarization I—Polarization by a Uniform Field. By R A. Buckingham Communicated by J E. Lennard-Jones, F.R S	94
The Quantum Theory of Atomic Polarization. II—The van der Waals Energy of Two Atoms. By R A Buckingham. Communicated by J. E. Lennard-Jones, F.R S	113
Equilibrium Curve and Entropy Difference between the Supraconductive and the Normal State in Pb, Hg, Sn, Ta, and Nb. By J G Daunt and K. Mendelssohn. Communicated by F. A. Landemann, F.R S	127
On Antiplane Stress in an Elastic Solid. By L. N. G. Filon, F R S	137

No. A 901—18 May 1937

On the Action of Ultra-Violet Sunlight upon the Upper Atmosphere. By M. N. Saha, F.R.S.	155
The Influence of Pressure on the Spontaneous Ignition of Inflammable Gas-Air Mixtures—The Sampler Olefines. By G. P. Kane and D. T. A. Townend. Communicated by W. A. Bone, F.R S	174

	PAGE
A Criticism of the Method of Expansion in Powers of the Gravitational Constant in General Relativity. By J. L. Synge. Communicated by E. T. Whittaker, F.R.S.	187
The Scattering of Alpha Particles in Helium, Hydrogen and Deuterium By C. B. O. Mohr and G. E. Pringle. Communicated by Lord Rutherford, O.M., F.R.S.	190
The Magneto-Resistance effect in Cadmium at Low Temperatures. By C. J. Milner. Communicated by Lord Rutherford, O.M., F.R.S.	207
Theory of Electrical Breakdown in Ionic Crystals. By H. Fröhlich. Communicated by N. F. Mott, F.R.S.	230
The Spectrum of Thallium Fluoride. By H. G. Howell. Communicated by W. E. Curtis, F.R.S. (Plate 1)	242
Adsorption on Measured Surfaces of Vitreous Silica—II. By W. G. Palmer. Communicated by Sir William Pope, F.R.S.	254
Properties of Sufficiency and Statistical Tests. By M. S. Bartlett. Communicated by R. H. Fowler, F.R.S.	268
The Accurate Determination of the Freezing-points of Alloys, and a Study of Valency Effects in Certain Alloys of Silver. By W. Hume-Rothery and P. W. Reynolds. Communicated by Sir Harold Carpenter, F.R.S.	282
The Energy Loss of Cosmic Ray Particles in Metal Plates. By P. M. S. Blackett, F.R.S. and J. G. Wilson	304

No. A 902—1 June 1937

On the Relation between Direct and Inverse Methods in Statistics. By Harold Jeffreys, F.R.S.	325
Series Evaluation of the Isotherm Data of CO ₂ between 0 and 150° C. and up to 3000 atm. By A. Michels and C. Michels. Communicated by F. A. Freeth, F.R.S.	348
The Isotherms of CO ₂ in the Neighbourhood of the Critical Point and Round the Coexistence Line. By A. Michels, B. Blaisse and C. Michels. Communicated by F. A. Freeth, F.R.S.	358

Contents

v

	PAGE
Thermodynamic Properties of CO ₂ up to 3000 Atmospheres between 25 and 150° C. By A. Michels, A. Bijl and C. Michels. Communicated by F. A. Freeth, F.R.S.	376
The Mercury Photosensitized Exchange Reaction of Deuterium and Phosphine. By H. W. Melville and J. L. Bolland. Communicated by James Kendall, F.R.S.	384
The Photochemical Decomposition and Oxidation of Trideuterophosphine. By H. W. Melville, J. L. Bolland and H. L. Roxburgh. Communicated by James Kendall, F.R.S.	406
On the Values of Fundamental Atomic Constants. By Sten von Friesen. Communicated by O. W. Richardson, F.R.S.	424
The Transport Numbers of the Ions in Solutions of Silver Dodecyl Sulphate. By O. Rhys Howell and Harry Warne. Communicated by Eric K. Rideal, F.R.S.	440
The Nuclear β -Rays of Radium D. By H. O. W. Richardson and Alice Leigh-Smith. Communicated by J. Chadwick, F.R.S. (Plates 2, 3)	454

No. A 903—15 June 1937

The Adsorption of Gases and the Equation of the Liquid State. By E. C. C. Baly, F.R.S.	465
The Band Systems ending on the $1s\sigma 2s\sigma^1\Sigma_g(^1X_g)$ state of H ₂ —Part I. By O. W. Richardson, F.R.S.	487
The Surface Layer of Polished Silica and Glass with Further Studies on Optical Contact. By Lord Rayleigh, F.R.S. (Plate 4)	507
The Photo-Electric Measurement of the Diurnal Variations in Daylight in Temperate and Tropical Regions. By W. R. G. Atkins, F.R.S., N. G. Ball and H. H. Poole	526
The Absorption Spectra and Photochemistry of Polyatomic Molecules Containing Alkyl Radicals. V—Vibration Frequencies and Structure. By H. W. Thompson and J. W. Linnett. Communicated by C. N. Hinshelwood, F.R.S.	539
The Ionization Potentials of the Free Radicals Methyl and Ethyl. By R. G. J. Fraser and T. N. Jewitt. Communicated by R. G. W. Norrish, F.R.S.	563

	PAGE
Physical Properties of Surfaces. IV—Polishing, Surface Flow and the Formation of the Beilby Layer. By F. P. Bowden and T. P. Hughes. Communicated by R. G. W. Norrish, F.R.S. (Plates 5–11) . . .	575
A Generalization of the Equations of the Self-Consistent Field for Two-Electron Configurations. By A. F. Stevenson. Communicated by D. R. Hartree, F.R.S.	588
Index	605

The Inverse Square Law of Gravitation—II

BY E. A. MILNE, F.R.S.

(Received 14 January 1937)

1—The general theory of relativity was largely developed owing to the failure of attempts to bring the law of gravitation within the scope of Lorentz transformations. Many such attempts were made in the early days of relativity. A notable one was contained in Poincaré's paper of 1906;* the astronomical consequences of this were developed by de Sitter in 1911. Other attempts are associated with the names of Abraham (1912),† Nordstrom (1912, 1913*a, b*), Behacker (1913) and others.

These attempts did not incorporate the phenomenon of the expanding universe, then undiscovered. Consequently they did not give recognition to the circumstance that, in accordance with Mach's principle, moving particles must be described in frames of reference associated with the actual distribution of matter-in-motion in the universe. By suitable graduation of the clocks carried by the fundamental observers associated with the frames defined by the extra-galactic nebular nuclei, these frames may be taken to be in uniform relative motion.‡ The present paper then obtains a completely relativistic formulation of the law of gravitation by purely kinematic methods.§ The results are expressed in the first instance in terms of t -measures, and in accordance with previous work they yield an inverse square law of gravitation with a "gravitational constant" γ proportional to the time t reckoned from the natural origin of time. When they are transformed to τ -measures,|| γ becomes a constant γ_0 , and the formulae are appropriate to classical or τ -dynamics. In t -measures we use the observers' flat private spaces; there proves to be no need to introduce any local curvature of space in the neighbourhood of a massive particle. In τ -measures we use a public hyperbolic static space in which the nebular nuclei appear at rest; the inverse square law of attraction then appears as

* For an alternative version of this work, see Cunningham (1914).

† See also his papers and discussions by Einstein (1912).

‡ The important theorem that this is always possible for a set of equivalent particles which can always signal to one another was proved by Whitrow (1935).

§ The preceding paper (Milne 1936*b*) obtained the inverse square law only in particular cases, and then only as an approximation. The present paper obtains an exact formulation holding generally.

|| The analysis embodying the τ -descriptions is given in a succeeding paper.

an approximation whose accurate expression is the hyperbolic function $\gamma_0 m_1 m_2 / (ct_0)^2 \sinh^2(\lambda/ct_0)$.

In this way the law of gravitation can be brought entirely within the scope of Lorentz formulae. The present investigation may be considered as continuing the investigation by Poincaré (1906).^{*} But whereas Poincaré assumed as a hypothesis that the gravitational potential must satisfy the wave equation, in the present *a priori* method the gravitational potential energy of two massive particles is deduced by kinematic methods outright, and it is verified *a posteriori* that it satisfies the wave-equation. This equation reduces to Laplace's equation in τ -measures, in agreement with empirical Newtonian theory. But the present method involves no empirical appeals, save to the existence of a temporal sequence in the experience of each individual observer, it proceeds without any specific theory of gravitation, or any adoption of "field equations".

2—In a previous series of papers the equation of motion of a particle P in the presence of the substratum or smoothed-out universe and under the action of an "external" field of potential χ has been obtained by kinematic methods. In another paper certain cases of the inverse square law of gravitation were derived, also by kinematic methods. The core of the present paper is the combination of these papers to obtain a general formula for the potential energy χ of any two massive particles m_1 and m_2 . This function χ has a remarkable set of properties which will be enumerated later. It is applied to obtain the relativistic equations of motion of a particle in the presence of a nebular nucleus, or of any two particles in one another's presence, taking into account the effect of the substratum. The consequences yield an explanation of the general spiral character of galaxies, and account for certain features of the motions observed in our own galaxy.

In the present investigation there is no "cosmical constant", nor any need for one.

DERIVATION OF A POTENTIAL-ENERGY FUNCTION

3—The general ideas and results developed in the previous paper under the same title† are supposed to be familiar to the reader. We proceed to extend these results so as to deal with the case of the interaction of any two massive particles in the presence of the substratum or smoothed-out universe. The notation is as before.

^{*} Poincaré wrote: "Je crois qu'il serait prématuré de vouloir pousser plus loin la discussion de ces formules."

† Milne (1936*b*). The other relevant references are Milne (1935*a*, 1936*a*, 1936*c* and 1937*a*, *b*).

Let O be a fundamental particle, taken as origin. It is a member of the *substratum*. Let P be any member of a *statistical distribution* or *universe*, as previously defined, described by a specified distribution-function $\psi(\xi)$. Then if P, V are the position-vector and velocity dP/dt of P at epoch t , in O 's reckoning, as measured by light-signalling on the kinematic scale, it has been shown by purely kinematic arguments that the acceleration of P is given by

$$\frac{dV}{dt} = -(\mathbf{P} - \mathbf{V}t) \frac{Y}{X} - (\mathbf{P} - \mathbf{V}t) \frac{Y}{X} \frac{C}{(\xi - 1)^{\frac{1}{2}}} \psi(\xi), \quad (1)$$

where

$$X = t^2 - \mathbf{P}^2/c^2, \quad Y = 1 - \mathbf{V}^2/c^2, \quad Z = t - \mathbf{P} \cdot \mathbf{V}/c^2, \quad \xi = Z^2/XY. \quad (2)$$

On the right-hand side of (1) the first term gives the acceleration in the presence of the substratum alone, and thus represents the "gravitational effect" of the substratum, and the second term represents the "effect" of the departure of the universe $\psi(\xi)$ from the smoothed-out state represented by the substratum.

4—We form as before a well-marked condensation by giving $\psi(\xi)$ a marked maximum at $\xi = 1$, and clear the inter-nebular space of matter by making $\psi(\xi)$ nearly zero everywhere else. For example we could take $\psi(\xi) = \psi(1)e^{-(\xi-1)^2/a^2}$, $a \sim 0$. Then the second term gives an inverse square attraction, locally towards the point Vt , for the local form of (1) is

$$\frac{dV}{dt} \sim -\frac{\mathbf{P} - \mathbf{V}t}{t^2} - \frac{\mathbf{P} - \mathbf{V}t}{|\mathbf{P} - \mathbf{V}t|^3} \frac{Cc^2t}{\psi(1)}. \quad (3)$$

Accordingly, as before, we define a number γ , which we call the "constant of gravitation", and a number m_2 , which we call the "gravitational mass" of the condensation, by the relations

$$\gamma m_2 = \frac{Cc^2t}{\psi(1)}, \quad (4)$$

$$\frac{4\pi}{3} \gamma m_0 \frac{B}{c^2 t^3} = \frac{1}{t^3}, \quad (5)$$

where m_0 is the arbitrary number assigned as the mass of a fundamental particle, the fundamental particles being distributed with particle-density

$$\frac{Bt \, dx \, dy \, dz}{c^3(t^2 - \mathbf{P}^2/c^2)^2}. \quad (6)$$

Definitions (4) and (5) are chosen so as to permit the description of the terms in (3) in the empirical Newtonian phraseology, though of course no appeal is made to the empirical law of gravitation. The factor $m_0 B/c^2 t^3$ in

(5) is the mass-density ρ_0 of the substratum at the origin O , at time t , as given by (6).

Solving (4) and (5) we have

$$\gamma = \frac{c^2 t}{M_0}, \quad (7)$$

$$m_2 = \frac{C}{\psi(1)} M_0, \quad (8)$$

where

$$M_0 = \frac{4}{3} \pi m_0 B. \quad (8')$$

Since $M_0 = \frac{4}{3} \pi (ct)^3 (m_0 B/c^2 t^3)$, M_0 may be interpreted as the mass of the fictitious universe* obtained by filling the accessible space at time t , namely a sphere of radius ct , homogeneously with matter of density ρ_0 .

For $\xi \sim 1$, (1) can now be written

$$\frac{dV}{dt} = -(\mathbf{P} - \mathbf{V}t) \frac{Y}{X} - (\mathbf{P} - \mathbf{V}t) \frac{Y}{X} \frac{m_2}{M_0(\xi - 1)^{\frac{1}{2}}}. \quad (9)$$

It is interpreted as giving the acceleration of P under the combined "pulls" of the whole substratum and of a condensation of mass m_2 located at the position $\mathbf{V}t$.

5—In any actual case, the precise degree of definiteness of the condensation will depend on the form assumed for $\psi(\xi)$. Different specifications of $\psi(\xi)$ will yield condensations of different degrees of sharpness. What we call a "massive particle" is a particular type of condensation, but hitherto the word "particle" has been just a useful nominative in the general theory, like "point" in a geometry.† We now define a "massive particle" or "gravitating particle" of mass m_2 as the type of condensation at $\mathbf{V}t$ for which (9) is exact. It will remain later to identify such a "gravitating particle" in the external world by properties derived from its definition and compared with observation.

6—Now let m_1 be the inertial mass assigned to the moving particle at P . Then by operations of which examples have been given in earlier papers, (9) may be written in the 4-vector form

$$\begin{aligned} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left[m_1 \xi^{\frac{1}{2}} \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right] \\ = -m_1 \xi^{\frac{1}{2}} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) \frac{1}{X} - \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) \frac{1}{X} \frac{m_1 m_2 \xi^{\frac{1}{2}}}{M_0(\xi - 1)^{\frac{1}{2}}} + \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m_1 \xi^{\frac{1}{2}}). \quad (10) \end{aligned}$$

* M_0 is the same as the actual mass assigned to the universe in certain cosmologies, namely 2.4×10^{40} g.

† It has been an element in the kinematic treatment capable of being described by a position-vector at epoch t , by means of light-signals and clock-readings.

It follows from (10) and from the definition of external force given in an earlier paper (Milne 1936*a*, § 43) that the particle m_1 moves as if acted on by an external force \mathbf{F} given by

$$\mathbf{F} = -\left(\mathbf{P} - \mathbf{V} \frac{Z}{Y}\right) \frac{1}{X} \frac{m_1 m_2 \xi^4}{M_0(\xi-1)^4} + \frac{\mathbf{V}}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4). \quad (11)$$

We now show that this force \mathbf{F} is derivable from a potential, as defined previously (Milne 1937*b*, § 7). For this will be so if we can find a function χ of \mathbf{P} and t such that

$$\mathbf{F} = -\frac{\partial \chi}{\partial \mathbf{P}} + 2 \frac{\mathbf{V}}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4). \quad (12)$$

Equating (11) and (12), we have to enquire whether a χ exists such that

$$\frac{\partial \chi}{\partial \mathbf{P}} = \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y}\right) \frac{1}{X} \frac{m_1 m_2 \xi^4}{M_0(\xi-1)^4} + \frac{\mathbf{V}}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4), \quad (13)$$

where $d(m_1 \xi^4)/dt$ is to be calculated from the equation of motion (10). To carry out this calculation we make use of the integrals of energy derived previously (Milne 1937*b*, § 4) for any equation of motion.

We associate with (11) a time-component F_t defined by

$$F_t = -\left(ct - c \frac{Z}{Y}\right) \frac{1}{X} \frac{m_1 m_2 \xi^4}{M_0(\xi-1)^4} + \frac{c}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4), \quad (11')$$

and with (12) the relation

$$F_t = + \frac{1}{c} \frac{\partial \chi}{\partial t} + 2 \frac{c}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4), \quad (12')$$

so that χ must also satisfy

$$\frac{1}{c} \frac{\partial \chi}{\partial t} = -\left(ct - c \frac{Z}{Y}\right) \frac{1}{X} \frac{m_1 m_2 \xi^4}{M_0(\xi-1)^4} - \frac{c}{Y^4} \frac{1}{Y^4} \frac{d}{dt} (m_1 \xi^4). \quad (13')$$

The energy-integral

$$F_t \frac{c}{Y^4} - \mathbf{F} \cdot \frac{\mathbf{V}}{Y^4} = \frac{1}{Y^4} \frac{d}{dt} (m_1 c^2 \xi^4) \quad (14)$$

is found to give only an identity, on using (11) and (11'). The energy-integral

$$\mathbf{F} \cdot \left(\frac{\mathbf{V}}{Y^4} - \mathbf{P} \frac{Y^4}{Z}\right) - F_t \left(\frac{c}{Y^4} - ct \frac{Y^4}{Z}\right) = \frac{1}{Y^4} \frac{d}{dt} (m_1 c^2 \xi^4) \quad (15)$$

will be found on insertion of (11) and (11') to yield

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m_1 \xi^{\frac{1}{2}}) = \frac{m_1 m_2}{M_0} \frac{1}{X^{\frac{1}{2}} (\xi - 1)^{\frac{1}{2}}}. \quad (16)$$

Inserting (16) in (13) we have to enquire whether a function χ exists such that

$$\begin{aligned} \frac{\partial \chi}{\partial \mathbf{P}} &= \frac{m_1 m_2 \xi^{\frac{1}{2}}}{M_0 (\xi - 1)^{\frac{1}{2}}} \left[\frac{\mathbf{P}}{X} - \mathbf{V} \frac{Z}{XY} + \frac{\mathbf{V}}{Z} \left(\frac{Z^2}{XY} - 1 \right) \right] \\ &= \frac{m_1 m_2}{M_0} \frac{\xi^{\frac{1}{2}}}{(\xi - 1)^{\frac{1}{2}}} \left[\frac{\mathbf{P}}{X} - \frac{\mathbf{V}}{Z} \right], \end{aligned} \quad (17)$$

and similarly such that at the same time

$$\frac{1}{c} \frac{\partial \chi}{\partial t} = \frac{m_1 m_2}{M_0} \frac{\xi^{\frac{1}{2}}}{(\xi - 1)^{\frac{1}{2}}} \left[-\frac{ct}{X} + \frac{c}{Z} \right]. \quad (17')$$

The four scalar relations implied by (17) and (17') are compatible, and determine χ as*

$$\chi = -\frac{m_1 m_2}{M_0} \frac{c^2 \xi^{\frac{1}{2}}}{(\xi - 1)^{\frac{1}{2}}}. \quad (18)$$

This is accordingly the potential of the field in which m_1 may be described as moving when it is in the presence of a massive particle m_2 at $(\mathbf{V}t, t)$. The sign of χ shows that the interaction is of the nature of an attraction.

7—It must now be noted that though ξ in (18) is a function of \mathbf{P} , \mathbf{V} and t , \mathbf{V} has been regarded merely as a parameter in deriving (18) from (17) and (17'); \mathbf{V} here simply defines the position of m_2 as $\mathbf{V}t$. Continuing to regard \mathbf{V} as a parameter, we find that χ as given by (18) satisfies identically the relation

$$\frac{\partial^2 \chi}{\partial x^2} + \frac{\partial^2 \chi}{\partial y^2} + \frac{\partial^2 \chi}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0, \quad (19)$$

where x, y, z are the Cartesian components of \mathbf{P} .

8—Written out *in extenso*, (18) is

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{t - \mathbf{P} \cdot \mathbf{V} / c^2}{[(t - \mathbf{P} \cdot \mathbf{V} / c^2)^2 - (t^2 - \mathbf{P}^2 / c^2)(1 - \mathbf{V}^2 / c^2)]^{\frac{1}{2}}}. \quad (20)$$

We now eliminate \mathbf{V} from this formula by introducing the position-vector of m_2 . We must take care to distinguish between epochs at m_1 and m_2 .

* This formula was suggested earlier (Milne 1936*b*, § 22), but in that context it was not derived *a priori* and was only shown to hold good locally.

Replace in (20) t by t_1 , \mathbf{P} by \mathbf{P}_1 . Then let t_2 denote an epoch at m_2 , so that if \mathbf{P}_2 is the position-vector of m_2 , $\mathbf{P}_2 = \mathbf{V}t_2$. Hence we replace \mathbf{V} in (20) by \mathbf{P}_2/t_2 . The result is

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2}{[(t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2)^2 - (t_1^2 - \mathbf{P}_1^2 / c^2)(t_2^2 - \mathbf{P}_2^2 / c^2)]^{1/2}}, \quad (21)$$

$$\text{or, say,} \quad \chi = -\frac{m_1 m_2 c^2}{M_0} \frac{X_{12}}{(X_{12}^2 - X_1 X_2)^{1/2}} = -\frac{m_1 m_2 c^2}{M_0} \frac{\zeta^{1/2}}{(\zeta - 1)^{1/2}}, \quad (22)$$

$$\text{where} \quad X_1 = t_1^2 - \mathbf{P}_1^2 / c^2, \quad X_2 = t_2^2 - \mathbf{P}_2^2 / c^2, \quad X_{12} = t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2,$$

$$\zeta = X_{12}^2 / X_1 X_2. \quad (23)$$

Formula (21) is to be interpreted in any actual application by setting $t_1 = t_2 = t$, where t is O 's reckoning of the epoch. As it stands, it is symmetrical in the suffixes 1 and 2, and it is a function only of the co-ordinates of events at P_1 and P_2 . By interchanging the roles of m_1 and m_2 , we see from the symmetry that inertial mass and gravitational mass are equal. We have been led to this formula by purely kinematic arguments for the particular case $\mathbf{P}_2 = \mathbf{V}t_2$, where \mathbf{V} is the momentary velocity of m_1 , but we have had no choice as to the way in which (20) was to be generalized so as to make it a function of t_1 , \mathbf{P}_1 , t_2 , \mathbf{P}_2 . We must therefore adopt (21) as giving the potential at P_1 due to P_2 , or at P_2 due to P_1 . It remains to enquire whether it yields the properties of gravitation. The properties it possesses are of course quite independent of its mode of derivation, and independent of any criticisms that the latter may be open to.

PROPERTIES OF THE GRAVITATIONAL POTENTIAL χ

§—We proceed to enumerate the properties of the function χ .

(1) *Invariance*—For given events (t_1, \mathbf{P}_1) , (t_2, \mathbf{P}_2) at m_1 and m_2 the function χ is of invariant value and invariant form for transformations from O to any other fundamental observer O' . For X_1 , X_2 , X_{12} are 4-scalar invariants under such transformations. Of course the interpretative relation $t_1 = t_2 = t$ is not invariant, events simultaneous for O are not simultaneous for O' . Hence the value of χ at O 's world-wide instant t is not equal to its value at O' 's world-wide instant t . The device of distinguishing between epochs at m_1 and m_2 in this context has been employed by other writers, but formula (21) does not seem to have been previously given.

(2) *Wave-equations*—Without further calculation it follows that χ satisfies the relations

$$\frac{\partial^2 \chi}{\partial x_1^2} + \frac{\partial^2 \chi}{\partial y_1^2} + \frac{\partial^2 \chi}{\partial z_1^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_1^2} = 0, \quad (24)$$

$$\frac{\partial^2 \chi}{\partial x_2^2} + \frac{\partial^2 \chi}{\partial y_2^2} + \frac{\partial^2 \chi}{\partial z_2^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_2^2} = 0. \quad (24')$$

The reader who verifies these in detail from (21) will find that their satisfaction depends essentially on the number of spatial dimensions being 3. Relations (24) and (24') are satisfied independently of setting $t_1 = t_2 = t$. It is particularly to be noted that we have not *postulated* the satisfaction of (24), (24'); that they are satisfied is an *a posteriori* property of the form of χ deduced kinematically.

It is of interest to note that if we had adopted some definition of a "massive particle" other than (9), for example by introducing on the right-hand side of (9) as a coefficient of m_2 some additional factor $\phi(\xi)$ such that $\phi(1) = 1$, we should have been able to deduce a potential function χ but it would not have satisfied a wave-equation. The satisfaction of (24) shows that at all points other than m_2 , the substratum has the properties usually assumed to be characteristic of "empty space", and so helps to identify our particular choice of condensation as representing an isolated particle.

(3) *Agreement with the empirical Newtonian potential*—By Lagrange's identity, χ may be written in the form

$$\chi = -\frac{m_1 m_2 c^3}{M_0} \frac{t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2}{[(t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1)^2 - c^{-2} (\mathbf{P}_1 \wedge \mathbf{P}_2)^2]^{1/2}}, \quad (25)$$

$$\text{whence} \quad \chi_{t_1=t_2=t} = -\frac{m_1 m_2 c^2 t}{M_0} \frac{1 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2 t^2}{[(\mathbf{P}_2 - \mathbf{P}_1)^2 - (\mathbf{P}_1 \wedge \mathbf{P}_2)^2 / c^2 t^2]^{1/2}}, \quad (26)$$

$$\text{or, by (7),} \quad \chi = -\gamma m_1 m_2 \frac{1 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2 t^2}{[(\mathbf{P}_2 - \mathbf{P}_1)^2 - (\mathbf{P}_1 \wedge \mathbf{P}_2)^2 / c^2 t^2]^{1/2}}. \quad (27)$$

When the origin O is chosen at m_1 , $\mathbf{P}_1 = 0$ and we have

$$\chi = -\frac{\gamma m_1 m_2}{|\mathbf{P}_2|}. \quad (28)$$

Similarly, when the origin O is chosen at m_2 , $\mathbf{P}_2 = 0$ and we have

$$\chi = -\frac{\gamma m_1 m_2}{|\mathbf{P}_1|}. \quad (28')$$

Again, when $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{P}_2| \ll ct_2$, we have for any origin not too far from m_1 and m_2 ,

$$\chi \sim -\frac{\gamma m_1 m_2}{|\mathbf{P}_2 - \mathbf{P}_1|}. \quad (28'')$$

Formulae (28), (28'), (28'') coincide with the ordinary Newtonian account of the potential and finally identify our condensation as a particle. But it must be remembered that here, in the t -dynamics, $\gamma \propto t$. The present value of γ has been shown previously to agree numerically with the observed Newtonian value (Milne 1936*a*, Appendix 1)

(4) *Limiting value of χ* —When O is such that the massive particle m_2 appears near the frontier of the observable universe in the t -description, we have $|\mathbf{P}_2| \sim ct_2$, $X_2 \sim 0$. We see that

$$\lim_{|\mathbf{P}_2| \rightarrow ct_2} \chi = -\frac{m_1 m_2 c^2}{M_0}, \quad (29)$$

a constant independent of \mathbf{P}_1 or t_1 . The same limit is found if $|\mathbf{P}_1| \rightarrow ct_1$. This non-zero limit replaces the Newtonian result which makes $\chi \rightarrow 0$ as $|\mathbf{P}_1|$ or $|\mathbf{P}_2| \rightarrow \infty$. The modification is consistent with the finite size of the expanding universe in t -measures.

THE INVERSE SQUARE LAW

10—By the general theory (Milne 1937*b*), the force \mathbf{F}_1 "acting on" m_1 due to m_2 is given by

$$\mathbf{F}_1 = -\frac{\partial \chi}{\partial \mathbf{P}_1} + 2 \frac{\mathbf{V}_1}{Y} \frac{1}{Y} \frac{d}{dt_1} (m_1 \xi \mathbf{t}). \quad (30)$$

Let us consider the component of this force represented by the gradient of χ . We find from (22)

$$-\frac{\partial \chi}{\partial \mathbf{P}_1} = \frac{m_1 m_2}{M_0} \frac{\xi^2}{(\xi - 1)^2} \left[\frac{\mathbf{P}_2}{X_{12}} - \frac{\mathbf{P}_1}{X_1} \right] \quad (31)$$

$$= \frac{m_1 m_2}{M_0} \frac{X_2}{(X_{12}^2 - X_1 X_2)^{3/2}} [\mathbf{P}_2 X_1 - \mathbf{P}_1 X_{12}] \quad (32)$$

$$= X_2 \frac{\partial}{\partial \mathbf{P}_2} \left[\frac{\chi}{X_{12}} \right]. \quad (33)$$

We make several deductions from these formulae.

(1) *The attraction*—Choose first the origin O to be at the attracting particle m_2 . Then $\mathbf{P}_2 = 0$, and (32) reduces to

$$\left(-\frac{\partial \chi}{\partial \mathbf{P}_1} \right)_{\mathbf{P}_2=0} = -\frac{m_1 m_2 c^2 t}{M_0} \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3} = -\gamma m_1 m_2 \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}. \quad (34)$$

This is a force acting on P_1 directed from P_1 towards P_2 of magnitude

$$\frac{\gamma m_1 m_2}{(\text{distance})^2}, \quad (35)$$

where the distance is measured by the fundamental observer coincident with the *attracting* particle. This appears therefore as the precise formulation of the Newtonian inverse square law.

Secondly, choose the origin O to be at the *attracted* particle m_1 . Then $\mathbf{P}_1 = 0$, and (32) reduces to

$$\left(-\frac{\partial \chi}{\partial \mathbf{P}_1}\right)_{\mathbf{P}_1=0} = \frac{m_1 m_2 c^2}{M_0} \frac{\mathbf{P}_2}{|\mathbf{P}_2|^3} \left(1 - \frac{\mathbf{P}_2^2}{c^2 t^2}\right) = \gamma m_1 m_2 \frac{\mathbf{P}_2}{|\mathbf{P}_2|^3} \left(1 - \frac{\mathbf{P}_2^2}{c^2 t^2}\right). \quad (36)$$

This is a force acting on P_1 directed from P_1 towards P_2 of the magnitude (35) times a correcting factor

$$1 - \frac{(\text{distance})^2}{(\text{radius of universe})^2}, \quad (35')$$

where now, in both (35) and (35') the distance is measured by the fundamental observer coincident with the *attracted* particle. The correcting factor is of course completely negligible in ordinary experience.

Since $-\partial \chi / \partial \mathbf{P}_1$ possesses these properties, it is convenient to call $-\partial \chi / \partial \mathbf{P}_1$ the "attraction" on m_1 .

(2) *Time-component*—The time-component of the attraction is given by

$$+\frac{1}{c} \frac{\partial \chi}{\partial t_1} = \frac{m_1 m_2 c}{M_0} \frac{\zeta^1}{(\zeta - 1)^{\frac{1}{2}}} \left[\frac{t_2}{X_{12}} - \frac{t_1}{X_1} \right] \quad (31')$$

$$= \frac{m_1 m_2 c}{M_0} \frac{X_2}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} [t_2 X_1 - t_1 X_{12}] \quad (32')$$

$$= -X_2 \frac{\partial}{c \partial t_1} \left[\frac{\chi}{X_{12}} \right]. \quad (33')$$

This vanishes when O is chosen to be at m_1 , ($\mathbf{P}_1 = 0$). It is negligibly small for $|\mathbf{P}_1|, |\mathbf{P}_2| \ll ct$, i.e. in ordinary experience.

(3) *Vanishing of attraction due to particle at boundary*—For any O , the attraction $-\partial \chi / \partial \mathbf{P}_1$ vanishes for $|\mathbf{P}_2| = ct_2$. Thus a particle at the confines of the observable t -universe exerts zero attraction at any internal point, although the t -measure of its distance is finite. It is readily verified that when the attraction vanishes, so also does the force \mathbf{F}_1 .

(4) *Newton's Third Law*—We have nowhere assumed the equality of action and reaction. We are now in a position to test it. We first notice that it is always approximately satisfied when the separation of the particles is

small compared with the radius of the universe. For the attraction on P_2 is given by

$$-\frac{\partial\chi}{\partial\mathbf{P}_2} = \frac{m_1 m_2}{M_0} \frac{X_1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} [\mathbf{P}_1 X_2 - \mathbf{P}_2 X_{12}], \quad (32a)$$

and so for $t_1 = t_2 = t$, in any frame, when $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{P}_2| \ll ct_2$, we have

$$-\frac{\partial\chi}{\partial\mathbf{P}_1} \sim \gamma m_1 m_2 \frac{\mathbf{P}_2 - \mathbf{P}_1}{|\mathbf{P}_2 - \mathbf{P}_1|^3}, \quad -\frac{\partial\chi}{\partial\mathbf{P}_2} \sim \gamma m_1 m_2 \frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|^3}. \quad (37)$$

It is, however, evident from comparison of (32) and (32a) that *measured in the same frame* the two attractions are not exactly equal and opposite. But if we measure the attractions in *corresponding* frames, say $-\partial\chi/\partial\mathbf{P}_1$ for O at P_2 and $-\partial\chi/\partial\mathbf{P}_2$ for O at P_1 , then the attractions are equal and opposite, by (34), whenever, at epochs t ,

$$|\mathbf{P}_1|_{t,-0} = |\mathbf{P}_2|_{t,-0}$$

This equality is satisfied whenever P_1 and P_2 are themselves fundamental particles, i.e. when they are equivalent. For then

$$|\mathbf{P}_1|_{t,-0} = |\mathbf{V}|t, \quad |\mathbf{P}_2|_{t,-0} = |\mathbf{V}|t,$$

where \mathbf{V} is the common value of the relative velocity. Thus, as we might have expected, equivalence carries with it the equality of action and reaction. The same result follows from (36) if we interchange origins

It is of interest to recognize that Newton's Third Law, in its various exact and approximate forms, is a deduction from our kinematic treatment, and is not needed as an independent postulate (as in Newton's formulation) or as a definition of equality of force (as in Mach's treatment). To many minds Newton's Third Law has appeared as the most mysterious of the laws of classical mechanics. The present analysis appears to throw some light on its origin, by relating its exact formulation to the idea of equivalence.

(5) *Analytical identities*—From (33) we have

$$-\frac{\partial\chi}{\partial\mathbf{P}_2} = X_1 \frac{\partial}{\partial\mathbf{P}_1} \left[\frac{\chi}{X_{12}} \right]. \quad (33a)$$

Differentiating (33) and (33a) we have the tensor relations

$$-\frac{1}{X_2} \frac{\partial^2 \chi}{\partial\mathbf{P}_1 \partial\mathbf{P}_1} = \frac{\partial^2}{\partial\mathbf{P}_1 \partial\mathbf{P}_2} \left[\frac{\chi}{X_{12}} \right] = -\frac{1}{X_1} \frac{\partial^2 \chi}{\partial\mathbf{P}_2 \partial\mathbf{P}_2}. \quad (38)$$

Contracting,

$$-X_1 \nabla_1^2 \chi = X_1 X_2 \frac{\partial^2}{\partial\mathbf{P}_1 \cdot \partial\mathbf{P}_2} \left[\frac{\chi}{X_{12}} \right] = -X_2 \nabla_2^2 \chi. \quad (39)$$

We find similarly

$$-X_1 \frac{\partial^2 \chi}{\partial t_1^2} = X_1 X_2 \frac{\partial^2}{\partial t_1 \partial t_2} \left[\frac{\chi}{X_{12}} \right] = -X_2 \frac{\partial^2 \chi}{\partial t_2^2}. \quad (40)$$

Again, by Euler's theorem on homogeneous functions,

$$t_1 \frac{\partial \chi}{\partial t_1} + \mathbf{P}_1 \cdot \frac{\partial \chi}{\partial \mathbf{P}_1} = 0, \quad t_2 \frac{\partial \chi}{\partial t_2} + \mathbf{P}_2 \cdot \frac{\partial \chi}{\partial \mathbf{P}_2} = 0. \quad (41)$$

We find also

$$+ \frac{1}{X_2} \left[t_2 \frac{\partial \chi}{\partial t_1} + \mathbf{P}_2 \cdot \frac{\partial \chi}{\partial \mathbf{P}_1} \right] = \frac{\chi}{X_{12}} = + \frac{1}{X_1} \left[t_1 \frac{\partial \chi}{\partial t_2} + \mathbf{P}_1 \cdot \frac{\partial \chi}{\partial \mathbf{P}_2} \right], \quad (42)$$

$$\mathbf{P}_1 \wedge \left(-\frac{\partial \chi}{\partial \mathbf{P}_1} \right) + \mathbf{P}_2 \wedge \left(-\frac{\partial \chi}{\partial \mathbf{P}_2} \right) = 0, \quad (43)$$

$$\frac{\mathbf{P}_2}{X_2} \wedge \left(-\frac{\partial \chi}{\partial \mathbf{P}_1} \right) + \frac{\mathbf{P}_1}{X_1} \wedge \left(-\frac{\partial \chi}{\partial \mathbf{P}_2} \right) = 0. \quad (44)$$

Identity (43) shows that in any frame the total moment about O of the mutual gravitational attractions on the system of the two particles is zero—in spite of the fact that the interactions are not *in general* along the line joining the two particles.

EQUATIONS OF MOTION UNDER GRAVITATIONAL FORCES

11—We have still to show that χ may be regarded as the mutual potential energy of the two particles. To do this we require their equations of motion. By the general theory developed (Milne 1937*b*, §§ 5, 7), the equations of motion are

$$\frac{1}{Y_1} \frac{d}{dt_1} \left[m_1 \xi_1 \frac{\mathbf{V}_1}{Y_1} \right] = -m_1 \xi_1 \left(\mathbf{P}_1 - \mathbf{V}_1 \frac{Z_1}{Y_1} \right) \frac{1}{X_1} - \frac{\partial \chi}{\partial \mathbf{P}_1} + 2 \frac{\mathbf{V}_1}{Y_1} \frac{1}{Y_1} \frac{d}{dt_1} (m_1 \xi_1), \quad (45)$$

$$\frac{1}{Y_1} \frac{d}{dt_1} \left[m_1 \xi_1 \frac{c}{Y_1} \right] = -m_1 \xi_1 \left(ct_1 - c \frac{Z_1}{Y_1} \right) \frac{1}{X_1} + \frac{1}{c} \frac{\partial \chi}{\partial t_1} + 2 \frac{c}{Y_1} \frac{1}{Y_1} \frac{d}{dt_1} (m_1 \xi_1), \quad (45')$$

$$\frac{1}{Y_2} \frac{d}{dt_2} \left[m_2 \xi_2 \frac{\mathbf{V}_2}{Y_2} \right] = -m_2 \xi_2 \left(\mathbf{P}_2 - \mathbf{V}_2 \frac{Z_2}{Y_2} \right) \frac{1}{X_2} - \frac{\partial \chi}{\partial \mathbf{P}_2} + 2 \frac{\mathbf{V}_2}{Y_2} \frac{1}{Y_2} \frac{d}{dt_2} (m_2 \xi_2), \quad (46)$$

$$\frac{1}{Y_2} \frac{d}{dt_2} \left[m_2 \xi_2 \frac{c}{Y_2} \right] = -m_2 \xi_2 \left(ct_2 - c \frac{Z_2}{Y_2} \right) \frac{1}{X_2} + \frac{1}{c} \frac{\partial \chi}{\partial t_2} + 2 \frac{c}{Y_2} \frac{1}{Y_2} \frac{d}{dt_2} (m_2 \xi_2), \quad (46')$$

where

$$\mathbf{V}_1 = d\mathbf{P}_1/dt_1, \quad \mathbf{V}_2 = d\mathbf{P}_2/dt_2.$$

Since $Y_1 dt_1$, $Y_2 dt_2$ are invariants for any events at P_1 , P_2 , for all observers O , and since only 4-vectors and 4-scalars involved, these equations of

motion are invariant in form for all fundamental observers O . For any given O , they are to be interpreted in his world-wide instant t by solving them with $t_1 = t_2 = t$, after the differentiations have been performed. Equations (45) and (45') are connected by one scalar identity, for on multiplying them respectively by \mathbf{P}_1 (scalarly) and ct_1 and subtracting we are left with identity (41). We have thus in effect two vector equations for determining $\mathbf{P}_1, \mathbf{P}_2$ as functions of t .

Each one of identities (42), (43), (44) may be used to eliminate χ from the equations of motion, and so yield a relation independent of the interaction. In particular (43) should yield a relation corresponding to the usual integral of angular momentum. Here we shall be content with obtaining the energy-integral.

12—Multiply (45) scalarly by \mathbf{V}_1 and (45') by c and subtract. Many terms cancel and we are left with

$$\frac{\partial \chi}{\partial t_1} + \mathbf{V}_1 \cdot \frac{\partial \chi}{\partial \mathbf{P}_1} + \frac{d}{dt_1} (m_1 c^2 \xi_1) = 0. \quad (47)$$

Similarly, from (46) and (46'),

$$\frac{\partial \chi}{\partial t_2} + \mathbf{V}_2 \cdot \frac{\partial \chi}{\partial \mathbf{P}_2} + \frac{d}{dt_2} (m_2 c^2 \xi_2) = 0. \quad (47')$$

$$\begin{aligned} \text{But} \quad \frac{d\chi}{dt} &= \left(\frac{\partial \chi}{\partial t_1} + \frac{\partial \chi}{\partial \mathbf{P}_1} \cdot \frac{d\mathbf{P}_1}{dt_1} \right) \frac{dt_1}{dt} + \left(\frac{\partial \chi}{\partial t_2} + \frac{\partial \chi}{\partial \mathbf{P}_2} \cdot \frac{d\mathbf{P}_2}{dt_2} \right) \frac{dt_2}{dt} \\ &= \left[\frac{\partial \chi}{\partial t_1} + \mathbf{V}_1 \cdot \frac{\partial \chi}{\partial \mathbf{P}_1} + \frac{\partial \chi}{\partial t_2} + \mathbf{V}_2 \cdot \frac{\partial \chi}{\partial \mathbf{P}_2} \right]_{t_1=t_2=t} \\ &= - \frac{d}{dt} [m_1 c^2 \xi_1 + m_2 c^2 \xi_2]_{t_1=t_2=t} \end{aligned}$$

by (47) and (47'). Hence, putting

$$E_1 = m_1 c^2 \xi_1 = M_1 c^2, \quad E_2 = m_2 c^2 \xi_2 = M_2 c^2,$$

$$\text{we have} \quad [\chi + E_1 + E_2]_{t_1=t_2=t} = \text{const.} \quad (48)^*$$

We see that the sum of invariants $(\chi + E_1 + E_2)_{t_1=t_2=t}$ remains constant in the experience of any observer measuring them. Since E_1 and E_2 have been identified (Milne 1936*a*, § 22) as the energies of the particles in motion, χ may be regarded as potential energy. The value of the constant depends on the observer O chosen.

* This is essentially different from Whittaker's integral (1935).

13—The intermediate equations (47), (47') are readily interpreted. Using (41) to eliminate $\partial\chi/\partial t_1$ we find from (47)

$$\frac{d}{dt_1}(m_1 c^2 \xi_1) = \left(-\frac{\partial\chi}{\partial \mathbf{P}_1}\right) \cdot \left(\mathbf{V}_1 - \frac{\mathbf{P}_1}{t_1}\right). \quad (49)$$

This exhibits the rate of increase of energy E_1 as equal to the rate of performance of work by the attraction $-\partial\chi/\partial \mathbf{P}_1$ in pushing the particle relative to its immediate cosmic environment, i.e. relative to the local standard of rest. If we insert (49) in (30) we complete the determination of external force in terms of χ .

14—Again, if we multiply (45') by \mathbf{V}_1/c and subtract it from (45), we obtain the actual acceleration of m_1 in the form

$$\frac{d\mathbf{V}_1}{dt_1} = -(\mathbf{P}_1 - \mathbf{V}_1 t_1) \frac{Y_1}{X_1} + \left(-\frac{\partial\chi}{\partial \mathbf{P}_1} - \frac{\mathbf{V}_1 \partial\chi}{c^2 \partial t_1}\right) \frac{Y_1}{m_1 \xi_1}. \quad (50)$$

Here, the first term on the right-hand side gives the component of acceleration due to the substratum alone. Accordingly the component of acceleration due to m_2 is

$$\left(-\frac{\partial\chi}{\partial \mathbf{P}_1} - \frac{\mathbf{V}_1 \partial\chi}{c^2 \partial t_1}\right) \frac{Y_1}{m_1 \xi_1}. \quad (51)$$

It will be noticed that distinct formulae are required for the "force", the "attraction" and the acceleration due to m_2 . We have successfully separated the "attraction" from the dynamical effects due to change of mass with velocity.

15—It would be natural to generalize the notion of gravitational potential to the case of any number of attracting particles by simple addition of the χ 's for each pair of particles. The sum-function would satisfy wave-equations and possess many of the properties already found. It would play the part of potential energy in the energy-integral of the aggregate of equations of motion. We could then proceed to deal with continuous density-distributions. Such developments will not be considered here.

16—Logically the next step is to transform the whole set of results now obtained into τ -measures, so as to make them appropriate to classical mechanics and the τ -scale of time. This is postponed to the succeeding paper. It will be more interesting to give now some examples of the theory in t -measures, although the full significance of the results will only be clear when we have the τ -forms for comparison.*

* It may be mentioned that the results about to be given were all obtained a long time ago, before the general (t, τ) transformation or the relativistic theory of gravitation had been fully developed. The particular case of the (t, τ) transformation for Keplerian orbits suggested the general transformation to the author, but this transformation was obtained from independent considerations by Leontovski and Nuut.

EXAMPLE. MOTION ROUND A NEBULAR NUCLEUS

17—Any given particle or star which is a member of a galaxy is moving under the combined action of the substratum or smoothed-out universe and the disturbances due to local aggregations of matter. In a preliminary investigation we can neglect the *disturbances* due to discrete external galaxies* (their combined effect is of course taken into account as the field of the substratum). We can further, to a first approximation neglect the material dispersed outside the nuclear region of the galaxy under consideration, and consider the mass of the galaxy as concentrated in its nucleus. We then have the simplest problem to which we can apply our analysis, namely the motion of a particle m in the neighbourhood of a large mass M located at a fundamental particle. For $M \gg m$, we can neglect the reaction of m on the nucleus, as is readily verified from equations (46), (46'). Without loss of generality we can now choose our origin O to be the fundamental observer at the nucleus M .

18—We apply the general theory by putting $m_1 = m$, $\mathbf{P}_1 = \mathbf{P}$, $m_2 = M$, $\mathbf{P}_2 = 0$. The potential χ reduces to

$$\chi = -\frac{mMc^3t}{M_0} \frac{1}{|\mathbf{P}|}, \quad (52)$$

where t is O 's measure of epochs at P . Then

$$-\frac{\partial \chi}{\partial \mathbf{P}} = -\frac{mMc^3t}{M_0} \frac{\mathbf{P}}{|\mathbf{P}|^3}, \quad + \frac{1}{c} \frac{\partial \chi}{\partial t} = -\frac{mMc^2}{M_0} \frac{1}{|\mathbf{P}|}. \quad (53)$$

Equations (45), (45') become, on removing a term from each side,

$$\frac{\xi^i}{Y^i} \frac{d}{dt} \left(\frac{V}{Y^i} \right) = -\xi^i \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) \frac{1}{X} - \frac{Mc^3t}{M_0} \frac{\mathbf{P}}{|\mathbf{P}|^3} + \frac{\mathbf{V}}{Y^i} \frac{1}{Y^i} \frac{d\xi^i}{dt}, \quad (54)$$

$$\frac{\xi^i}{Y^i} \frac{d}{dt} \left(\frac{c}{Y^i} \right) = -\xi^i \left(ct - c \frac{Z}{Y} \right) \frac{1}{X} - \frac{Mc^2}{M_0} \frac{1}{|\mathbf{P}|} + \frac{c}{Y^i} \frac{1}{Y^i} \frac{d\xi^i}{dt}. \quad (54')$$

These are the general relativistic equations for motion round a galactic nucleus M at O , as observed by an observer at O . In accordance with the general theory (54') is a consequence of (54); it can be deduced from (54) by scalar multiplication of (54) by \mathbf{P} .

19—By the usual procedure, the energy integral is readily found to be

$$c^2 \xi^i - \frac{Mc^3t}{M_0} \frac{1}{|\mathbf{P}|} = \text{const.} \quad (55)$$

* Although, as Jeans has suggested, the "tidal forces" due to them may be responsible for determining certain features of the forms of galaxies, such as the existence of a pair of spiral arms.

20—To find the integral of angular momentum, multiply (54) vectorially by \mathbf{P} . We find

$$\frac{\xi^i}{Y^i} \frac{d}{dt} \left[\frac{\mathbf{P} \wedge \mathbf{V}}{Y^i} \right] = \frac{\xi}{X^i} \frac{\mathbf{P} \wedge \mathbf{V}}{Y^i} + \frac{\mathbf{P} \wedge \mathbf{V}}{Y^i} \frac{1}{Y^i} \frac{d\xi^i}{dt}.$$

Using the identity
$$\frac{1}{Y^i} \frac{dX^i}{dt} \equiv \xi^i,$$

we find that the preceding equation may be rearranged as

$$X^i \frac{d}{dX^i} \left[\frac{1}{\xi^i} \frac{\mathbf{P} \wedge \mathbf{V}}{Y^i} \right] = \frac{1}{\xi^i} \frac{\mathbf{P} \wedge \mathbf{V}}{Y^i}.$$

This integrates in the form

$$\frac{1}{\xi^i} \frac{\mathbf{P} \wedge \mathbf{V}}{Y^i} = \mathbf{A} X^i,$$

or*

$$\mathbf{P} \wedge \mathbf{V} = \mathbf{A} Z. \quad (56)$$

The integral (56) shows that the radius vector \mathbf{P} is always perpendicular to a fixed vector \mathbf{A} , and therefore moves in a plane. An obvious generalization of (56), taking into account the angular momentum of the whole galaxy, would establish the general planar character of galaxies. Further (56) gives the usual secular increase of angular momentum with time t obtained in earlier papers. It is essentially different from the integral of angular momentum in the Newtonian or τ -dynamics, and proves to be one of the sources of the generally spiral character of resolved galaxies. The origin of the secular increase of angular momentum is of course the couple due to the substratum.

21—Now choose plane polar co-ordinates r, θ in the fixed plane in which P moves. Then

$$X = t^2 - r^2/c^2, \quad Y = 1 - (\dot{r}^2 + r^2\dot{\theta}^2)/c^2, \quad Z = t - r\dot{r}/c^2,$$

$$|\mathbf{P} \wedge \mathbf{V}| = r^2\dot{\theta}.$$

The integrals (55) and (56) are then sufficient to determine r and θ as functions of t .

22—We are primarily interested in the solution of equation (54) for particles P not too far from the nucleus and moving with speeds not comparable with c . Such particles will include all proper members of the galaxy. We can then neglect terms in $1/c^2$, but we retain of course the terms due to

* It will be noticed that in this example the angular momentum integral takes a form slightly different from that in certain earlier theorems. This is due to the specific form of the "external" gravitational force in this case.

the substratum. Under these circumstances, the r -equation of motion obtained from (54) reduces to

$$\ddot{r} - r\dot{\theta}^2 = -\frac{r - t\ddot{r}}{t^3} - \frac{t}{t_0} \frac{\gamma_0 M}{r^2}, \quad (57)$$

where we have put*

$$\gamma = \frac{c^3 t}{M_0} = \frac{\gamma_0 t}{t_0},$$

so that

$$\gamma_0 = c^3 t_0 / M_0. \quad (58)$$

The θ -equation can be replaced by the integral of angular momentum, (56), which reduces to

$$r^2 \dot{\theta} = h_0 \frac{t}{t_0}, \quad (59)$$

say. Equations (57) and (59) determine the motion of a particle near a nebular nucleus, with neglect of terms in $1/c^2$.

23—To solve (57) and (59), choose a new radial co-ordinate ρ defined by

$$r = \rho \frac{t}{t_0}, \quad (60)$$

and a new time-variable τ defined by

$$\frac{dt}{t} = \frac{d\tau}{t_0}, \quad \tau = t_0 \log(t/t_0) + t_0. \quad (61)$$

Then (57) and (59) reduce to

$$\frac{d^2 \rho}{d\tau^2} - \rho \left(\frac{d\theta}{d\tau} \right)^2 = -\frac{\gamma_0 M}{\rho^2}, \quad (62)$$

$$\rho^2 \frac{d\theta}{d\tau} = h_0. \quad (63)$$

24—Comparing (62) with (57) we see that the transformations (60) and (61) have removed the term representing the pull of the substratum and removed the secular dependence of the "gravitation-constant" on t . And comparison of (63) with (59) shows that in terms of ρ and τ angular momentum is now a constant. Equations (62) and (63) are in fact precisely the equations for a Keplerian orbit under a Newtonian attraction $\gamma_0 M/\rho^2$, provided ρ could be deemed to be the "real" radius-vector, τ the "real" time. It has been shown in preceding papers that when the clocks of all fundamental observers are regraduated from t to τ according to (61), then τ can be identified as the independent time-variable of classical dynamics, and ρ is the measure of the radial distance on the τ -scale of light-signalling.

* Here t_0 is the present value of t , and γ_0 (a constant) the present value of γ .

Thus the present *a priori* derivation of the law of gravitation and the equations of motion reduce to the empirical Newtonian formulation (in this instance) once we recognize τ as our ordinary time-variable. In the succeeding paper we shall establish this quite generally in its relativistic form.

25—For the time being we require only the general nature of the orbits in the t -measures. Equations (62) and (63) give of course elliptic, parabolic or hyperbolic orbits in τ -measures, and the transforms of the elliptic orbits to t -measures will give the orbits of proper members of the galaxy. Clearly an ellipse transforms into a spiral of radius increasing (on the average) with t , by (60).

An observer in the plane of the orbit, situated at the nucleus and using classical dynamics and τ -measures, will regard the orbit as an ellipse. But an observer using t -measures will regard the orbit as a spiral. And *a spiral is the form the orbit must exhibit to an observer viewing it from outside its plane*. For directions are unaltered by transformation from t to τ , and so an actual spiral path, if the whole of it could be viewed simultaneously, must appear as a spiral whether the observer is using t -measure of time or τ -measure. Actually we do not know whether the spiral coils of nebulae represent orbits or not. We only know that we are seeing a set of particles or stars strung out along their orbits. But our analysis seems to indicate the origin of the spiral character of resolved galaxies. If the outside observer is located on some other nebular nucleus, in t -measures the two nebular nuclei will be separating at a rate proportional to their distance, and a single orbit, of radius-vector from its own nucleus systematically increasing with the time, would have for its projection perpendicular to the line of sight an ellipse again. But if a set of particles or stars were strung out at different positions along one and the same orbit, we should in effect be viewing the whole orbit at one time, and the appearance would be that of a spiral.

Clearly great care will be required to disentangle all these effects. We may be content at this stage with our general tentative explanation of the spiral character of resolved nebulae as connected with the fact that the orbits being described are spirals when reckoned in t -measure, which is the measure in terms of which the nebulae are receding.

26—The elucidation of further details can best be dealt with in an astronomical paper, but it is worth while here sketching the resulting theory for one particular type of orbit, namely that corresponding in τ -measure to a circle. The outstanding characteristic of the internal motions of most celestial systems is that of *rotation*. Rotation has played such a great part

in theoretical cosmogony that it is appropriate to consider here briefly the t -analogue of rotation.

27—Rotation in τ -measure is just the orbit*

$$\rho = \text{const.} = \rho_0. \quad (64)$$

We wish to determine the t -description of this motion. By (60) the present distance r_0 , i.e. at $t = t_0$, of the particle undergoing (64) is given by

$$r_0 = \rho_0. \quad (65)$$

Introducing (64) in (62) and (63) we have the Keplerian formulae

$$\rho_0 \left(\frac{d\theta}{d\tau} \right)^2 = \frac{\gamma_0 M}{\rho_0^2}, \quad \rho_0^2 \frac{d\theta}{d\tau} = h_0, \quad (66)$$

$$\text{Hence} \quad h_0^2 = \gamma_0 M \rho_0. \quad (67)$$

Integrating either of (66), the position angle θ at time τ is given by

$$\theta - \theta_0 = \left(\frac{\gamma_0 M}{\rho_0^3} \right)^{1/2} (\tau - \tau_0),$$

where θ_0 is the present value of θ . Hence, by (61),

$$\theta - \theta_0 = \left(\frac{\gamma_0 M}{\rho_0^3} \right)^{1/2} t_0 \log \frac{t}{t_0}. \quad (68)$$

$$\text{But by (60)} \quad r = \rho_0 \frac{t}{t_0}. \quad (69)$$

Equations (68) and (69) give the orbit parametrically in terms of t . Eliminating t we have

$$\theta - \theta_0 = \left(\frac{\gamma_0 M t_0^2}{\rho_0^3} \right)^{1/2} \log \frac{r}{\rho_0},$$

$$\text{or, by (65),} \quad r = r_0 e^{(\theta - \theta_0) \cot \alpha_0}, \quad (70)$$

$$\text{where} \quad \tan \alpha_0 = \left(\frac{\gamma_0 M t_0^2}{r_0^3} \right)^{1/2}. \quad (71)$$

Orbit (70) is an equiangular spiral of angle α_0 . It is well known that the arms of spiral nebulae are approximately equiangular spirals.

28—Let us apply (71) to our own galaxy. For the mass interior to the sun, we may take, following J. S. Plaskett (1935, p. 22)†

$$M = 1.65 \times 10^{11} \odot = 3.3 \times 10^{44} \text{ g.}$$

* No confusion need be caused by our previous use of the symbol ρ_0 to denote the mean density of the substratum near the observer.

† It may be recalled that this datum was obtained by an application of Kepler's Third Law to galactic rotation. Since galactic rotation is determined spectroscopically, the velocities derived must be kinematic or t -velocities, but as we shall see in a moment the spiral orbits are here very nearly circular.

and the sun's distance from the galactic centre as

$$r_0 = 10^4 \text{ parsecs} = 3 \times 10^{23} \text{ cm.}$$

With $\gamma_0 = 6.66 \times 10^{-8} \text{ cm.}^3 \text{ sec.}^{-2} \text{ g.}^{-1}$, $t_0 = 0.6 \times 10^{17} \text{ sec.}$, we find from (71)

$$\tan \alpha_0 = 54, \quad \alpha_0 = 89^\circ.$$

This shows that at the sun's distance the spiral motion reduces to almost pure rotation. For a galaxy for which α_0 could be measured at a known linear distance r_0 , (71) could be used to determine the mass M .

29—We can also determine the velocities for the case considered. Let u_r , u_θ be the components of velocity in the galactic plane along and perpendicular to the radius from the galactic centre. Then

$$u_r = \dot{r} = \frac{d}{dt} \left(\rho_0 \frac{t}{t_0} \right) = \frac{\rho_0}{t_0}, \quad (72)$$

and
$$u_\theta = r\dot{\theta} = \rho_0 \frac{t}{t_0} \frac{d\theta}{dt} = \rho_0 \frac{d\theta}{d\tau} = \frac{h_0}{\rho_0} = \left(\frac{\gamma_0 M}{\rho_0} \right)^{\frac{1}{2}}. \quad (73)$$

Formula (72) gives the galactic K -effect. Formula (73) agrees with the result of applying Kepler's Third Law.

At the sun's distance from the centre, with the same data, (72) and (73) give

$$u_r = 5 \times 10^4 \text{ cm. sec.}^{-1} = 5 \text{ km. sec.}^{-1},$$

$$u_\theta = 2.7 \times 10^7 \text{ cm. sec.}^{-1} = 270 \text{ km. sec.}^{-1}.$$

The last result is of course merely recovery of the datum which led to Plaskett's value of M . The outward velocity u_r , since it is expressed in t -measure, should be apparent as a red-shift or K -term. The K -effect is well known, and has recently been confirmed by Smart (1936), and an amount of 5 km. sec.^{-1} is compatible with the observations.

If we eliminate ρ_0 between (72) and (73) we obtain

$$u_r = \frac{\gamma_0 M}{t_0 u_\theta^2}. \quad (74)$$

This determines the K -effect from the mass and the rotational velocity.

30—To find the stream-lines corresponding to the equiangular spiral orbits we write down u_r and u_θ in terms of r and t , thus

$$u_r = \frac{r}{t}, \quad u_\theta = \left(\frac{\gamma_0 M}{r} \frac{t}{t_0} \right)^{\frac{1}{2}}, \quad (75)$$

and integrate the equation
$$\frac{dr}{u_r} = \frac{r d\theta}{u_\theta},$$

keeping t constant. The result is

$$\theta - \theta_1 = -\frac{2}{3} \left(\frac{\gamma_0 M t^3}{t_0} \right)^{\frac{1}{3}} \frac{1}{r^{\frac{1}{3}}}. \quad (76)$$

At time $t = t_0$, the stream-lines are accordingly given by

$$\theta - \theta_1 = -\frac{2}{3} \left(\frac{\gamma_0 M t_0^3}{r^3} \right)^{\frac{1}{3}}. \quad (77)$$

These are spiral curves in the same sense as the orbits (r increasing with θ). The velocity of rotation of a stream-line as a whole is

$$\left(\frac{\partial \theta}{\partial t} \right)_{r=\text{const}} = - \left(\frac{\gamma_0 M t}{t_0} \right)^{\frac{1}{3}} \frac{1}{r^{\frac{1}{3}}}, \quad (77a)$$

and is accordingly retrograde. The value of $\partial \theta / \partial t$ is equal to $-u_\theta / r$, which is numerically the angular velocity in the orbit but of opposite sign.

The fact that the stream-lines are spirals is likely, again, to give a general spiral appearance to the nebula. The stream-lines have the property that $\theta \rightarrow$ finite limit as $r \rightarrow \infty$. Thus the stream-lines will become appreciably straight when

$$r^3 \gg \gamma_0 M t_0^3,$$

which gives for our own galaxy

$$r \gg 1.4 \times 10^4 \text{ parsecs.}$$

Our own galactic system seems to be confined within this size.

31.—We make one last application of our formulae. Suppose that in the vicinity of a given star S , say the sun, the motion is predominantly of the type we have been considering. Then the stars S' in the vicinity of galactic latitude b , galactic longitude l (measured from the prolongation of OS) will possess certain constants of differential motion relative* to S . The radial velocity observed from S is known to be given by†

$$SS' \cos^2 b \left[\frac{1}{2}(A+B) + \frac{1}{2}(A-B) \cos 2l + H \sin 2l \right], \quad (78)$$

and the cross-component of velocity parallel to the galactic plane is given by†

$$SS' \cos b [\omega_2 + H \cos 2l + \frac{1}{2}(A-B) \sin 2l], \quad (79)$$

* The peculiar motion of S is supposed to have been removed.

† These formulae are due essentially to Ört (1927). In the form quoted they have been derived by the author (1935b), but substantially the same formulae had been given earlier by Pilowski (1931) and by Ogorodnikov (1932).

$$\text{where} \quad \frac{1}{2}(A+B) = \frac{1}{2} \left[\frac{\partial u_r}{\partial r} + \frac{u_r}{r} + \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} \right], \quad (80)$$

$$\frac{1}{2}(A-B) = \frac{1}{2} \left[\frac{\partial u_r}{\partial r} - \frac{u_r}{r} - \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} \right], \quad (80')$$

$$H = \frac{1}{2} \left[\frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r} + \frac{1}{r} \frac{\partial u_r}{\partial \theta} \right], \quad (80'')$$

$$\omega_3 = \frac{1}{2} \left[\frac{\partial u_\theta}{\partial r} + \frac{u_\theta}{r} - \frac{1}{r} \frac{\partial u_r}{\partial \theta} \right]. \quad (80''')$$

When the motion is given by (75), the constants have the values

$$\frac{1}{2}(A+B) = \frac{1}{t}, \quad \frac{1}{2}(A-B) = 0, \quad (81)$$

$$H = -\frac{3}{4} \left(\frac{\gamma_0 M t}{r l_0} \right)^{\frac{1}{2}} \frac{1}{r} = -\frac{1}{4} \Omega, \quad \omega_3 = \frac{1}{4} \left(\frac{\gamma_0 M t}{r l_0} \right)^{\frac{1}{2}} \frac{1}{r} = \frac{1}{4} \Omega \quad (82)^*$$

where $\Omega = \dot{\theta} = u_\theta/r$. The result of interest is that the coefficient of $\cos 2l$ in (78) is zero. It follows that the maximum radial velocities occur in the directions $l = \frac{1}{4}\pi$, $l = \frac{3}{4}\pi$, and the minimum radial velocities in the directions $l = \frac{3}{4}\pi$, $l = \frac{1}{4}\pi$, hence one bisector of the angles between these directions points exactly to the galactic centre, $l = \pi$. This is the observed direction.

The coincidence of this direction with the galactic centre has been previously interpreted as meaning an absence of K -effect or galactic expansion; for, on pure rotation, ($u_r = 0$) we would have both $\frac{1}{2}(A+B) = 0$ and $\frac{1}{2}(A-B) = 0$. Result (81) shows that the observed coincidence of the bisector with the direction of the galactic centre is compatible with the existence of the galactic K -term predicted by the theory, $\frac{1}{2}(A+B)$. The present analysis determines the sign of the K -term as positive, and since $\frac{1}{2}(A+B)$ is independent of r , the differential K -term should be strictly proportional to SS' .

SUMMARY

32—By purely kinematic methods, the law of gravitation is obtained in relativistic form in the flat space of any fundamental observer using the t -dynamics. An expression is found for the mutual potential energy χ of any two massive particles. It is Lorentz-invariant, and it reduces under appropriate choice of observer to the Newtonian potential energy with a "constant of gravitation" proportional to the epoch t . It is used to calculate the "attraction", force and acceleration experienced by a particle m_1 in

* It is a general result that $\omega_3 - H = \Omega$.

the presence of a particle m_2 . The attraction is shown to be accurately describable by an inverse square law, provided the distance is measured by the fundamental observer coinciding with the attracting particle. The conditions of validity of the law of equality of action and reaction are obtained. The equations of motion of a pair of particles are obtained, and applied to the problem of the orbit of a particle in the presence of a nebular nucleus. It is shown that the orbits are equiangular spirals in the t -dynamics but appear as Keplerian orbits in the τ -dynamics, and the generally spiral character of resolved galaxies is accounted for. Sundry velocity-features of the motions observed in our own galaxy are explained.

REFERENCES

- Abraham 1912 *Phys. Z.* **13**, 4 and 311
 Behaeker 1913 *Phys. Z.* **14**, 989
 Cunningham 1914 "The Principle of Relativity," 1st ed. Cambridge.
 de Sitter 1911 *Mon. Not. R. Astr. Soc.* **71**, 388.
 Einstein 1912 *Ann. Phys., Lpz.* **38**, 39.
 Milne, E. A. 1935a "Relativity, Gravitation and World-Structure" Oxford
 — 1935b *Mon. Not. R. Astr. Soc.* **95**, 570
 — 1936a *Proc. Roy. Soc. A*, **154**, 22
 — 1936b *Proc. Roy. Soc. A*, **156**, 62
 — 1936c *Proc. Roy. Soc. A*, **158**, 324
 — 1937a *Proc. Roy. Soc. A*, **159**, 171.
 — 1937b *Proc. Roy. Soc. A*, **159**, 528
 Nordstrom 1912 *Phys. Z.* **13**, 1126
 — 1913a *Ann. Phys., Lpz.* **40**, 856
 — 1913b *Ann. Phys., Lpz.* **42**, 533
 Ogorodnikoff 1932 *Z. Astrophys.* **4**, 190
 Oort 1927 *Bull. Astr. Insts. Netherlands* **3**, 275.
 Pilowski 1931 *Z. Astrophys.* **3**, 53.
 Plaskett, J. S. 1935 Halley Lecture "The Structure of the Galaxy" Oxford.
 Poincaré 1906 *R.C. Circ. mat. Palermo*, **21**, 166.
 Smart 1936 *Mon. Not. R. Astr. Soc.* **96**, 588.
 Whitrow 1935 *Quart. J. Math.* **6**, 257.
 Whittaker, E. T. 1935 *Proc. Roy. Soc. A*, **149**, 390
-

The Inverse Square Law of Gravitation—III

By E. A. MILNE, F.R.S.

(Received 14 January 1937)

1—In a preceding paper a relativistic formulation of the law of gravitation was obtained, in the flat private space of any fundamental observer in the substratum or smoothed-out universe, in terms of t -measures. In other papers,* general formulae have been given for the transformation of forces, equations of motion, etc., from t -measures to τ -measures. They may be at once applied to express the law of gravitation in τ -measures, in the public hyperbolic space $d\epsilon^3$ in which the extra-galactic nebular nuclei appear at rest. The present paper carries out this programme, and so obtains the law of gravitation in the form appropriate to the τ -dynamics, which corresponds to classical mechanics.

2—Let O be any fundamental particle of the substratum, \mathbf{P} the position vector of any other particle P with respect to O , at epoch t , in t -measure. The t -clocks of the fundamental observers have been graduated so that the fundamental particles appear in uniform (Whitrow 1935) relative motion.

These clocks are then regraduated from t to τ , where

$$\tau = t_0 \log(t/t_0) + t_0, \quad (1)$$

t_0 being the present value of t . All measures must then be transformed accordingly. Let $\mathbf{P} = l\mathbf{i}$, where \mathbf{i} is a unit vector in the direction of polar co-ordinates θ, ϕ . Then it has been shown that the event (l, θ, ϕ, t) in t -measures is described in τ -measures as $(\lambda, \theta, \phi, \tau)$, where

$$t = t_0 e^{\frac{\tau-t_0}{t_0}} \cosh \frac{\lambda}{ct_0}, \quad (2)$$

$$l/c = t_0 e^{\frac{\tau-t_0}{t_0}} \sinh \frac{\lambda}{ct_0}. \quad (3)$$

* The relevant references are: "World-structure" Milne (1935); "On the foundations of dynamics" (Milne 1936a); "The inverse square law of gravitation" (Milne 1936b); "Kinematics, dynamics and the scale of time" (Milne 1936c, 1937a, b); "The inverse square law of gravitation—II", see p. 1.

3—Now consider two massive particles P_1, P_2 , of masses m_1, m_2 . It has been shown that associated with events (P_1, t_1) (P_2, t_2) at P_1 and P_2 there is a gravitational potential energy χ given by

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{X_{12}}{(X_{12}^2 - X_1 X_2)^{1/2}}, \quad (4)$$

$$\text{where } X_1 = t_1^2 - \mathbf{P}_1^2/c^2, \quad X_2 = t_2^2 - \mathbf{P}_2^2/c^2, \quad X_{12} = t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2/c^2, \quad (5)$$

and M_0 is the mass of the fictitious universe. In the experience of O , at O 's world-wide instant t , (4) is to be interpreted by putting $t_1 = t_2 = t$. The value of the constant of gravitation is

$$\gamma = c^2 t / M_0. \quad (6)$$

To transform (4) (which, being an energy, is a time-invariant) we put

$$\mathbf{P}_1 = l_1 \mathbf{i}_1, \quad \mathbf{P}_2 = l_2 \mathbf{i}_2, \quad (7)$$

and apply (2) and (3). We find in τ -measure

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{\cosh \frac{\lambda_1}{ct_0} \cosh \frac{\lambda_2}{ct_0} - \sinh \frac{\lambda_1}{ct_0} \sinh \frac{\lambda_2}{ct_0} (l_1 \cdot l_2)}{\left[\left(\cosh \frac{\lambda_1}{ct_0} \cosh \frac{\lambda_2}{ct_0} - \sinh \frac{\lambda_1}{ct_0} \sinh \frac{\lambda_2}{ct_0} (l_1 \cdot l_2) \right)^2 - 1 \right]^{1/2}}. \quad (8)$$

This may be written in the alternative form

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \times \frac{\cosh \frac{\lambda_1}{ct_0} \cosh \frac{\lambda_2}{ct_0} - \sinh \frac{\lambda_1}{ct_0} \sinh \frac{\lambda_2}{ct_0} (l_1 \cdot l_2)}{\left[\left(\cosh \frac{\lambda_1}{ct_0} \sinh \frac{\lambda_2}{ct_0} l_1 - \cosh \frac{\lambda_2}{ct_0} \sinh \frac{\lambda_1}{ct_0} l_1 \right)^2 - \sinh^2 \frac{\lambda_1}{ct_0} \sinh^2 \frac{\lambda_2}{ct_0} (l_1 \wedge l_2)^2 \right]^{1/2}}. \quad (8')$$

Formula (8) or (8') gives the potential energy of two particles of masses m_1, m_2 , at positions $(\lambda_1, l_1), (\lambda_2, l_2)$ respectively.

4—It will be noticed that in (8) and (8') mention of the epochs τ_1 and τ_2 has completely disappeared; χ is now a function of the spatial co-ordinates of P_1 and P_2 only. This is a particular case of a general theorem given previously (Milne 1937b, § 15), according to which any potential function $\chi(\mathbf{P}, t)$ when transformed to τ -measures becomes explicitly independent of τ . It follows that in τ -measure it is not necessary to specify the epoch to which χ refers.

5.—To see the significance of (8) and (8') choose the origin O to be on the line joining P_1P_2 , in the sense OP_1P_2 . Then $\lambda_2 > \lambda_1$, and $l_1 = l_2$, so that $l_1 \cdot l_2 = 1$. Formula (8) now becomes

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{1}{\tanh \frac{\lambda_2 - \lambda_1}{ct_0}}. \quad (9)$$

This may be written

$$\chi = -\frac{\gamma_0 m_1 m_2}{ct_0 \tanh \frac{\lambda_2 - \lambda_1}{ct_0}}, \quad (9')$$

where γ_0 is the present value of γ , given by

$$\gamma_0 = c^2 t_0 / M_0. \quad (6')$$

The numerical value of γ_0 , deduced from the recession of the galaxies and the mean density of matter in space near ourselves, has been shown to be about equal to the empirical value of the Newtonian constant of gravitation (Milne 1936*a*, Appendix 1).

Formula (9') gives the mutual potential energy of m_1 at P_1 and m_2 at P_2 in the public static hyperbolic space

$$de^2 = d\lambda^2 + (ct_0)^2 \sinh^2 \frac{\lambda}{ct_0} (d\theta^2 + \sin^2 \theta d\phi^2).$$

Since for collinear particles co-ordinate distances λ have been shown (Milne 1937*a*, § 22) to be additive, $(\lambda_2 - \lambda_1)$ is independent of the origin O in the line P_1P_2 , and χ depends only on the τ -measure of the distance between the particles.

For $(\lambda_2 - \lambda_1) \ll ct_0$, (9') reduces to

$$\chi \sim -\frac{\gamma_0 m_1 m_2}{\lambda_2 - \lambda_1}. \quad (10)$$

This agrees with the empirical Newtonian expression for the potential energy. We have seen in the preceding paper that in t -measure the Newtonian formula is exact; we now see that in τ -measure the exact formula for χ involves mention of ct_0 , the present radius of the universe in t -measure. In ordinary experience it is immaterial whether we use τ -measure or t -measure.

6.—The private Euclidean spaces of the t -observers extend only to distance ct , but the public hyperbolic space de^2 extends to infinity. Formula (9) shows that as $(\lambda_2 - \lambda_1) \rightarrow \infty$,

$$\chi \rightarrow -\frac{m_1 m_2 c^2}{M_0}. \quad (11)$$

We have already had a similar result in t -measures. This result must be contrasted with the limit of the empirical formula (10), which is zero. The change is connected with the circumstance that it is impossible to remove a particle "to a large distance from all attracting matter", for a particle is always in the presence of the substratum.

7—The "attraction" exerted by P_2 on P_1 has been defined as $-\text{grad } \chi$. In τ -measure, when O is chosen collinear with P_1 and P_2 , the λ -component of this attraction is

$$-\frac{\partial \chi}{\partial \lambda_1} = \frac{\gamma_0 m_1 m_2}{(ct_0)^2 \sinh^2 \frac{\lambda_2 - \lambda_1}{ct_0}}. \quad (12)$$

This is directed towards P_2 . For $(\lambda_2 - \lambda_1) \ll ct_0$, it gives

$$-\frac{\partial \chi}{\partial \lambda_1} \sim \frac{\gamma_0 m_1 m_2}{(\lambda_2 - \lambda_1)^2}. \quad (13)$$

This is the inverse square law of attraction, of which (12) appears to be the exact form in τ -measure. The attraction tends to zero as $(\lambda_2 - \lambda_1) \rightarrow \infty$. Further, when O is collinear with P_1 and P_2 ,

$$\left(-\frac{\partial \chi}{\partial \lambda_1}\right) + \left(-\frac{\partial \chi}{\partial \lambda_2}\right) = 0, \quad (14)$$

so that action and reaction are here equal and opposite.

8—In the previous paper it was shown that χ in t -measure satisfies the differential equations

$$\nabla_1^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_1^2} = 0, \quad \nabla_2^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_2^2} = 0. \quad (15)$$

The τ -measure of χ must therefore satisfy the τ -transformations of (15)

Now if to the wave equation

$$\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0 \quad (16)$$

we apply the transformation of co-ordinates (2), (3), we find that a factor $e^{-\frac{1}{2} \frac{\tau^2 - t^2}{t_0^2}}$ divides through, and we are left with

$$\frac{1}{\sinh^2 \frac{\lambda}{ct_0}} \frac{\partial}{\partial \lambda} \left[\sinh^2 \frac{\lambda}{ct_0} \frac{\partial \chi}{\partial \lambda} \right] + \frac{1}{(ct_0)^2 \sinh^2 \frac{\lambda}{ct_0}} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \chi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \chi}{\partial \phi^2} \right] - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial \tau^2} - \frac{2}{ct_0} \frac{\partial \chi}{\partial \tau} = 0. \quad (17)$$

Since χ in τ -measure is explicitly independent of τ_1 and τ_2 , it follows that χ as given by (8) satisfies the equations

$$\frac{\partial}{\partial \lambda_1} \left[(ct_0)^3 \sinh^2 \frac{\lambda_1}{ct_0} \frac{\partial \chi}{\partial \lambda_1} \right] + \frac{1}{\sin \theta_1} \frac{\partial}{\partial \theta_1} \left(\sin \theta_1 \frac{\partial \chi}{\partial \theta_1} \right) + \frac{1}{\sin^2 \theta_1} \frac{\partial^2 \chi}{\partial \phi_1^2} = 0, \quad (18)$$

$$\frac{\partial}{\partial \lambda_2} \left[(ct_0)^3 \sinh^2 \frac{\lambda_2}{ct_0} \frac{\partial \chi}{\partial \lambda_2} \right] + \frac{1}{\sin \theta_2} \frac{\partial}{\partial \theta_2} \left(\sin \theta_2 \frac{\partial \chi}{\partial \theta_2} \right) + \frac{1}{\sin^2 \theta_2} \frac{\partial^2 \chi}{\partial \phi_2^2} = 0, \quad (18')$$

where we have put

$$\mathbf{i}_1 = (\cos \theta_1, \sin \theta_1 \cos \phi_1, \sin \theta_1 \sin \phi_1),$$

$$\mathbf{i}_2 = (\cos \theta_2, \sin \theta_2 \cos \phi_2, \sin \theta_2 \sin \phi_2).$$

Equations (18), (18') are just Laplace's equation in the space $d\epsilon^3$. It will be seen that the dual description of gravitation by means of t -measures and τ -measures removes the old difficulties associated with the *propagation* of gravitation. A gravitational potential will appear to be propagated as waves moving with the velocity of light in t -measure, but the same potential will appear as propagated instantaneously in τ -measure.

As a verification, take the case when O is at P_1 . Then $\lambda_1 = 0$, and χ reduces to

$$\chi = -\frac{\gamma_0 m_1 m_2}{ct_0 \tanh \frac{\lambda_2}{ct_0}}. \quad (19)$$

Since it is now explicitly independent of θ_2 and ϕ_2 , (18') reduces to

$$\frac{\partial}{\partial \lambda_2} \left[(ct_0)^3 \sinh^2 \frac{\lambda_2}{ct_0} \frac{\partial}{\partial \lambda_2} \frac{1}{ct_0 \tanh \frac{\lambda_2}{ct_0}} \right] = 0,$$

which is obviously identically satisfied. For $\lambda_1, \lambda_2 \ll ct_0$, (18) and (18') reduce to Laplace's equation in flat space in spherical polars.

A complete theory of the potential in the hyperbolic space $d\epsilon^3$ could now be constructed from the elementary solution (19). We shall not undertake this here.

9—In accordance with the general theory given previously, the force on P_1 exceeds the attraction on P_1 by a term involving the rate of change of mass of P_1 . We have (Milne 1937*b*, § 15, equation (45))

$$\phi_{\lambda_1} = -\frac{\partial \chi}{\partial \lambda_1} + 2 \frac{\partial \mathcal{F}_1}{\partial \lambda_1} \frac{1}{(1 - 2\mathcal{F}_1/m_1 c^2)} \frac{d}{d\tau} \left[\frac{1}{(1 - 2\mathcal{F}_1/m_1 c^2)^{1/2}} \right], \quad (20)$$

$$\text{where} \quad \mathcal{F}_1 = \frac{1}{2} m_1 \left[\dot{\lambda}_1^2 + (ct_0)^2 \sinh^2 \frac{\lambda_1}{ct_0} (\dot{\theta}_1^2 + \sin^2 \theta_1 \dot{\phi}_1^2) \right], \quad (21)$$

dots denoting differentiations with respect to τ . We have similar formulae for Φ_{θ_1} and Φ_{ϕ_1} .

When P_1 is in the vicinity of O , ($\lambda_1 \ll ct_0$), we can put $\Pi_1 = \lambda_1 i_1$, and write the force in the local vector form

$$\Phi_1 = -\frac{\partial \chi}{\partial \Pi_1} + \frac{2v_1}{(1-v_1^2/c^2)} \frac{d}{d\tau} \left[\frac{m_1}{(1-v_1^2/c^2)} \right], \quad (22)$$

where $v_1 = d\Pi_1/d\tau$, $\chi = -\gamma_0 m_1 m_2 / |\Pi_2 - \Pi_1|$.

10—Again, by the general theory (Milne 1937*b*, § 15, equation (46)), the equations of motion of P_1 under the "action" of P_2 and the substratum are

$$(1 - 2\mathcal{F}_1/m_1 c^2)^{-1} \left[\frac{d}{d\tau} \left(\frac{\partial \mathcal{F}}{\partial \dot{\lambda}_1} \right) - \frac{\partial \mathcal{F}}{\partial \lambda_1} \right] = -\frac{\partial \chi}{\partial \lambda_1}, \quad (23)$$

and two similar equations in θ_1 and ϕ_1 . There is a similar set of three equations of motion for P_2 . Combining the six with multipliers $\dot{\lambda}_1$, $\dot{\theta}_1$, etc., we readily find the energy integral

$$\frac{m_1 c^2}{(1 - 2\mathcal{F}_1/m_1 c^2)^{1/2}} + \frac{m_2 c^2}{(1 - 2\mathcal{F}_2/m_2 c^2)^{1/2}} + \chi = \text{const.} \quad (24)$$

Equations (23), etc., are the general equations of motion for the two-body problem, in τ -measures, for its description by any observer at "local rest", i.e. by any observer sharing the mean motion of the nearby nebular nuclei.

11—In our ordinary experience the two particles P_1 and P_2 will be at distances from a fundamental particle not comparable with ct_0 (which is 2×10^9 light-years). If we take $\lambda_1, \lambda_2 \ll ct_0$, the space becomes locally Euclidean and the equations of type (23) may be put in the vector forms

$$\left(1 - \frac{v_1^2}{c^2}\right)^{-1} m_1 \frac{dv_1}{d\tau} = -\frac{\partial \chi}{\partial \Pi_1} = -\gamma_0 m_1 m_2 \frac{\Pi_1 - \Pi_2}{|\Pi_1 - \Pi_2|^3}, \quad (25)$$

$$\left(1 - \frac{v_2^2}{c^2}\right)^{-1} m_2 \frac{dv_2}{d\tau} = -\frac{\partial \chi}{\partial \Pi_2} = -\gamma_0 m_1 m_2 \frac{\Pi_2 - \Pi_1}{|\Pi_2 - \Pi_1|^3}. \quad (25')$$

In these equations the velocities v_1, v_2 are not restricted to be small compared with c , but they must be measured in the frame at local cosmical rest.

Equations (25), (25') possess the energy integral

$$\frac{m_1 c^2}{(1 - v_1^2/c^2)^{1/2}} + \frac{m_2 c^2}{(1 - v_2^2/c^2)^{1/2}} - \frac{\gamma_0 m_1 m_2}{|\Pi_1 - \Pi_2|} = \text{const.} \quad (26)$$

We can eliminate the interaction terms in (25) and (25') by multiplying them vectorially by Π_1 and Π_2 , respectively, and adding, as in the usual procedure

for obtaining the angular momentum integral, but the resulting equation does not appear to be integrable. We find in fact

$$\frac{m_1 \Pi_1 \wedge d\mathbf{v}_1/d\tau}{(1 - \mathbf{v}_1^2/c^2)^{3/2}} + \frac{m_2 \Pi_2 \wedge d\mathbf{v}_2/d\tau}{(1 - \mathbf{v}_2^2/c^2)^{3/2}} = 0. \quad (27)$$

When the velocities are small compared with c , (27) integrates in the form

$$\Pi_1 \wedge m_1 \mathbf{v}_1 + \Pi_2 \wedge m_2 \mathbf{v}_2 = \text{const.},$$

which is the usual angular momentum integral. The motion is then in a plane, and is the ordinary Keplerian motion.

12—We cannot however use (25) and (25') to calculate the *relative* orbit of either particle *as observed from the other*, in general, for we do not know the transformations from a fundamental particle-observer to an observer in motion relative to him, in the τ -dynamics. To a first approximation such transformations will be the transformations of Newtonian relativity, but the exact transformations may be expected to differ from these by terms in $1/c^2$, for the finding of which no method has yet been given. Thus (25) and (25'), though they should yield the motion of a planet round the sun *as observed from the centre of the galaxy* (or in the mean velocity frame of the galaxies in our own neighbourhood), are not competent to yield the relativistic refinements of the relative motion of sun and planet *as observed from either sun or planet*, for the sun is not at local cosmical rest—it is, roughly speaking, rotating round the galactic centre besides having a peculiar motion of its own. The sun's velocity relative to the standard of local rest may be expected to be an essential datum in the calculation of these refinements, for example in the calculation of the motion of perihelion of a planet. Equations (25) and (25'), in τ -measures, are invariant in form for transformations from any frame at local nebular rest to any other such frame, but they must not be expected to be invariant for transformations to other frames, even to frames in uniform relative motion with respect to "local rest". The reason is that whilst all fundamental particles are "equivalent" in the technical sense, a fundamental particle is not equivalent to a non-fundamental particle such as the sun. In the frame associated with a non-fundamental particle, i.e. in terms of observations carried out by an observer not at local nebular rest, the relation to the distribution of matter-in-motion in the universe will be different, and accordingly the derived laws of dynamics, which (in agreement with Mach's conjectures) we have shown to be intimately connected with the distribution of matter-in-motion in the universe, will be slightly different. Thus, although (25) and (25') embody a rigorous formulation of the general two-body problem in

τ -measures, as it would appear to an observer on a nebular nucleus or other particle at local rest, they require transformation in a manner not yet discovered before they can be used to yield the orbit to any other type of observer, for example an observer on either particle. It is much to be hoped that some method will be evolved for determining such transformations.* In the meantime we may be content with having determined the equations of the two-body problem in frames at local rest, and with pointing out the possibility of deviations from the motions predicted by Newtonian or "general" relativity consequent on one of the bodies concerned not being at local rest. It is possible that such effects as the unexplained terms in the moon's motion which are outside current gravitational theory, may be due to the earth's or sun's motion in the frame defined by the local standard of rest.

13—Since (25) and (25') do not appear to admit an exact integral of angular momentum, the relative orbit does not appear to lie accurately in a plane. In certain special cases we can however obtain an integral of angular momentum. These will now be considered.

14—Let one of the particles be itself a fundamental particle, namely a nebular nucleus of large mass M . Then the reaction on it of the second particle may be neglected. Put $m_2 = M$, $m_1 = m$, and take $\Pi_2 = 0$, $\Pi_1 = \Pi$. Then Π_2 , by (25'), will be permanently zero if $m \ll M$, and we have the single equation

$$(1 - v^2/c^2)^{-1} \frac{d\mathbf{v}}{d\tau} = -\gamma_0 M \frac{\Pi}{|\Pi|^3}, \quad (28)$$

The energy integral is now

$$\frac{c^2}{(1 - v^2/c^2)^{1/2}} - \frac{\gamma_0 M}{|\Pi|} = \text{const.}, \quad (29)$$

and by multiplying (28) vectorially by Π we obtain

$$\Pi \wedge \frac{d\mathbf{v}}{d\tau} = 0,$$

which integrates to yield the integral of angular momentum

$$\Pi \wedge \mathbf{v} = \text{const.} = \mathbf{h}_0. \quad (30)$$

This is Kepler's Law of Areas in the τ -dynamics. The orbit of m , judged from O , now lies in a plane perpendicular to \mathbf{h}_0 . That (30) is consistent with

* Since the above was written, the required transformations have been found

the corresponding integral of angular momentum in t -measure as previously (previous paper, formula (56)) obtained, namely

$$\mathbf{P} \wedge \mathbf{V} = \mathbf{A}Z, \quad (31)$$

may be seen as follows. Writing $Z = X^\dagger Y^\dagger \xi^\dagger$, transforming to τ -measure by using

$$\xi^\dagger = d\tau/d\sigma, \quad Y^\dagger dt = ds = e^{\frac{\tau-t_0}{t_0}} d\sigma, \quad X^\dagger = t_0 e^{\frac{\tau-t_0}{t_0}},$$

the integral (31) becomes

$$ct_0 e^{\frac{\tau-t_0}{t_0}} \sinh \frac{\lambda}{ct_0} \frac{1}{e^{\frac{\tau-t_0}{t_0}}} \frac{d\tau}{d\sigma} \frac{d}{d\tau} \left[ct_0 e^{\frac{\tau-t_0}{t_0}} \sinh \frac{\lambda}{ct_0} 1 \right] = \mathbf{A} \frac{d\tau}{d\sigma} t_0 e^{\frac{\tau-t_0}{t_0}},$$

$$\text{or, for } \lambda \ll ct_0, \quad \Pi = \lambda 1, \quad \Pi \wedge \frac{d\Pi}{d\tau} = \mathbf{A} t_0, \quad (32)$$

which is (30). We see once again how the secular increase of angular momentum in t -measure given by (31) reduces to a constant angular momentum in τ -measure.

The integrals (29) and (30) may now be expressed in plane polar co-ordinates and yield a Keplerian orbit in τ -measure modified by terms in $1/c^2$. The description of this orbit in t -measure has been considered in detail in Part II (§§ 17–27) of the papers under the present title, and shown to explain the general spiral character of motion round a nebular nucleus.

15—Again, let one of the two particles be a non-fundamental particle (e.g. the sun) of mass M large compared with the other particle m . Then, by (25'), \mathbf{v}_2 , the velocity of M , is approximately constant. Hence we can write (25) in the form

$$(1 - \mathbf{v}_1^2/c^2)^{-1} \frac{d^2}{d\tau^2} [\Pi_1 - \Pi_2] = -\gamma_0 M \frac{\Pi_1 - \Pi_2}{|\Pi_1 - \Pi_2|^3}. \quad (33)$$

This admits the integral of angular momentum

$$(\Pi_1 - \Pi_2) \wedge \frac{d}{d\tau} (\Pi_1 - \Pi_2) = \text{const.} = \mathbf{h}_0, \quad (34)$$

say, and the relative vector $(\Pi_1 - \Pi_2)$, as viewed by a neighbouring fundamental particle O , lies in a fixed plane perpendicular to \mathbf{h}_0 . The energy integral reduces to

$$\frac{c^2}{(1 - \mathbf{v}_1^2/c^2)^{1/2}} - \frac{\gamma_0 M}{|\Pi_1 - \Pi_2|} = \text{const.} \quad (35)$$

Equation (33) or the integral (35) shows that even the relative orbit, as viewed by O , depends to terms in $1/c^2$ on the absolute velocity \mathbf{v}_1 of the small

particle m in the frame at local rest. This is of course an effect of the substratum, since v_1 is the velocity relative to the substratum in τ -measure. The further solution of (33) lies more in the province of an astronomical paper.

RECAPITULATION

16—The theory of dynamics and gravitation which has been developed in the present series of papers should be competent to deal with the general problems of the motions in galaxies, the proper motions of galaxies, the effects (tidal or otherwise) of galaxies on one another and any problem involving the relations of moving matter to the general distribution of matter in the whole universe. The theory is a purely logical one, reposing on the existence of a temporal sequence in the experience of each individual observer (Milne 1935, chap. 2). The theorems which have been deduced, which appear as laws of dynamics and gravitation, are purely logical consequences of the definition of a substratum as a set of "equivalent particles", equivalent both in the kinematical ($A \equiv B$) and statistical ($A \equiv B$) senses. An early theorem (Milne 1935, §§ 96–103) showed that, without any appeal to empirical or assumed laws of dynamics or gravitation, such a system will continue *of itself*, without the imposition of external constraints, in the prescribed motion. The system was thus shown to be a *natural* one, and it was shown by Whitrow that any system of equivalent particles, whatever their relative accelerations as judged by time-signalling conducted with the clocks belonging to them, could be reduced to a substratum in uniform relative motion by appropriate regraduation of these clocks. The author had in fact previously obtained the general transformations for the accelerated motion of equivalent particles, of which particular cases are the classical Lorentz formulae (Milne 1935, chap. 2, equations (33)).

A fundamental step was then the determination of the *free* motion of an additional test-particle in the presence of a substratum. This was embodied in two theorems, first the determination of the form of the acceleration formula in a form involving a certain function $G(\xi)$, secondly* the determination of $G(\xi)$ for a substratum as $G(\xi) \equiv -1$, independent of the particle-density in the substratum. This permitted the fixation of the measure of external force F acting on a *constrained* particle in terms of the departure of the actual motion from the calculated free motion, and in turn the construction of a complete dynamics of such constrained motions. Rate of performance of work was defined as the complete scalar product of a force

* A purely kinematical proof of this result is given by Milne (1937c).

into the velocity of the particle relative to its immediate environment in the substratum. This led to the identification of $mc^2\dot{\xi}^4$, a function of the position, velocity and epoch of the particle considered, as *energy*.^{*} The dynamical properties of the substratum could be described as those of a conservative "gravitational" field in which the constant of gravitation was taken to be c^2/M_0 .

At this stage the resulting dynamics, though formally resembling classical dynamics in certain relations, such as Einstein's relation $E = Mc^2$, appeared as essentially different from it. The measures of the fundamental dynamical magnitudes involved specification of the position and epoch of a particle, relative to an observer O , as well as of its velocity, but more surprisingly there appeared in equations of motion a term representing the gravitational pull of the substratum. It was then shown that the dynamical magnitudes reduced to their classical expressions, and that the cosmical acceleration-term in the equations of motion completely disappeared, if the clocks of all fundamental particles, previously graduated to read the time t in which the substratum or universe was expanding uniformly, were regraduated to read time τ , where

$$\tau = t_0 \log(t/t_0) + t_0,$$

t_0 being the present value of t . This regraduation of clocks carried with it an adjustment of all derived magnitudes to the new time-scale τ . Certain magnitudes, called "time-invariants", proved to be unaltered by the regraduation—energy was an example—but others had different measures. For example, γ reduced to a constant $c^2 t_0/M_0$, and angular momentum, shown to increase proportionally to t in t -measure, reduced (for a semi-isolated system) to a constant in τ -measure; "velocity relative to environment" in t -measure reduced simply to "velocity" in τ -measure. In all cases the τ -measures played precisely the same parts in the transformed equations as their classical counterparts play in classical equations. The conclusion was inferred that in classical mechanics the independent time-variable† is τ , not t . Moreover τ , the epoch-coordinate of any event (whether at the observer or not), proved to be an invariant, the same for all fundamental observers and consequently enjoying all the properties of a world-wide Newtonian time. The resulting *local* equations of motion differ slightly from Einstein's "special relativity" mechanics but possess the same energy integral

^{*} The formula for energy was shown by Whitrow (1936) to be applicable to photons.

† It had long been realized that "general" relativity, in treating all co-ordinates as on the same footing, is not in a position to identify the independent variable representing the time, as was explicitly recognized by De Sitter.

It thus appears that the time t is appropriate in some contexts, the time τ in others. Dynamics has a dual structure. In τ -measures, appropriate where we use ordinary dynamics and gravitation, the fundamental particles appear as relatively stationary, the universe appears as a non-expanding system, and the natural origin of time $t = 0$ is relegated to $\tau = -\infty$. This removes certain difficulties concerning the time-scale of astronomy. The secular increase of angular momentum in t -measure explains the general prevalence of angular momentum in isolated "pockets" in the universe generally. It is convenient in τ -measures to use a public static hyperbolic space in which to describe motion and position, whereas in t -measures it is convenient to use observers' private flat spaces. In the hyperbolic space the motion of a free particle obeys the Galileo-Newton principle of inertia, which replaces the accelerations due to the substratum appearing in the t -dynamics. It follows that inertia is simply the pull of the rest of the universe.

The red-shifts observed in the spectra of extra-galactic nebulae indicate that atoms keep the time t , not τ . The universe appears as finite, of radius ct , in t -measures, but infinitely extending in τ -measures. Thus the old questions of whether the universe is or is not expanding, whether it is finite or infinite in extent, are differently answered according to the scale of time and dynamics adopted. The results obtained in these papers are intimately connected with the refusal to adopt the notion of a rigid body as an indefinable concept; instead, length-measures are defined in terms of light-time, whose numerical value depends on the scale of clock-graduation adopted. A rod rigid on the τ -scale is deemed to be uniformly expanding on the t -scale.

Concurrently with the general dynamics, an account of gravitation has been developed, without any empirical appeal to quantitative laws of nature or adoption of assumed "field equations". It was first shown how to define a conservative field of potential in t -measures. Then, by generalizing the notion of a substratum so as to make it contain condensations and by reducing these condensations to isolated massive particles, we determined the potential energy to be associated with a pair of massive particles in such a way that the actual motion of each, determined from kinematical considerations, corresponded to the "field" of the other. This was expressed in t -measure, in flat space, in Lorentz-invariant form. This determined the law of gravitation by kinematic methods, and led to its exact formulation in inverse square form. It was applied to the motion of a particle under the influence of the substratum and a neighbouring galactic nucleus. The combination of the secular increase in angular momentum with the secular increase in γ led to general spiral motions reducing to Keplerian motions in τ -measures.

This was of course a particular case of the general result that kinematic arguments lead to a dynamics and gravitation (expressed in t -measures) which reduce to classical mechanics and gravitation in τ -measures. Descriptions in t -measures seem the more fundamental, as they explicitly involve the dependence of the phenomena on epoch t ; in descriptions in τ -measures, the present epoch t_0 appears as a constant in exact formulae, though disappearing in our ordinary approximations, but τ -descriptions are essentially ephemeral. Every phenomenon may equally be described with the same logical correctness in either mode of time-reckoning, but it is important to know which mode is being employed in any given case.

SUMMARY

17—The law of gravitation for any two particles of masses m_1 and m_2 , as derived in a preceding paper, is expressed in τ -measure, i.e. in a form appropriate to the public hyperbolic space $d\epsilon^3$ in which the system of nebular nuclei appears as at relative rest. For two particles separated by co-ordinate distance λ , the potential energy χ is given by

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{1}{\tanh \frac{\lambda}{ct_0}} = -\frac{\gamma_0 m_1 m_2}{ct_0 \tanh \frac{\lambda}{ct_0}},$$

where γ_0 is a constant $c^2 t_0 / M_0$ equal to the Newtonian constant of gravitation. This reduces to the Newtonian expression for $\lambda \ll ct_0$. χ satisfies Laplace's equation in hyperbolic space, and yields the inverse square law of attraction with the appropriate modification for hyperbolic space. The relativistic form of the equations of motion for the two-body problem is obtained, and certain cases are discussed. The present account of dynamics and gravitation, constructed by purely kinematic methods, is recapitulated.

REFERENCES

- Milne, E. A. 1935 "Relativity, Gravitation and World-Structure." Oxford
 — 1936a *Proc. Roy. Soc. A*, **154**, 22.
 — 1936b *Proc. Roy. Soc. A*, **156**, 62.
 — 1936c *Proc. Roy. Soc. A*, **158**, 324.
 — 1937a *Proc. Roy. Soc. A*, **159**, 171.
 — 1937b *Proc. Roy. Soc. A*, **159**, 526.
 — 1937c *Quart. J. Math.* **8**, 22.
 Whitrow 1935 *Quart. J. Math.* **6**, 257.
 — 1936 *Quart. J. Math.* **7**, 271.

A Theoretical Formula for the Solubility of Hydrogen in Metals

By R. H. FOWLER, F.R.S. AND C. J. SMITHELLS, D.Sc.

(Received 26 January 1937)

1—In certain metals such as Cu, hydrogen appears to be dissolved in the metal in the form of free protons, which do not affect the normal metal lattice, even when present at very considerable concentrations. In other metals such as Ti, definite metal hydrides are formed which have a different lattice structure from the pure metal. The metal Pd is intermediate since the hydrogen affects the lattice constant. It is the properties of the former group of metals which are first to be discussed here, since the fact that the normal metal lattice is (practically) unaffected seems to justify a very simple theoretical treatment of the solubility, and it is of some interest to examine how the theory compares with the facts. We shall find that we can bring the facts and the theory into satisfactory order together. The various types of solubility curve are shown in fig. 1.

2—From evidence such as the well-known $p^{1/2}$ law for the rate of diffusion of hydrogen through metals we may certainly assume that the hydrogen in the metal is atomic. For the present we shall neglect the difference between atoms of hydrogen and protons plus electrons, and merely assume that the atoms are present as such in the metal, without specific interaction with particular metallic atoms; the metal merely provides a region in which hydrogen atoms can exist and move in a definite field of potential energy. Specific contributions by the electrons of the hydrogen atoms will be considered later, when the hydrogen atoms in the metal will be considered as protons plus electrons.

The problem is now sufficiently precise. We have hydrogen atoms distributed between two enclosures or phases (i) outside the metal where they are present as atoms (very few in number) and molecules (in the main) in a field of potential energy zero, (ii) inside the metal where they are present in a field of potential energy $\chi_e + W$, say (χ_e constant). The regions where $\chi_e + W$ has its least value may be assumed, in view of the absence of specific interactions, and lattice changes, to be fairly extensive, and we shall choose W so that $W = 0$ in these regions while elsewhere $W > 0$, χ_e thus represents the excess energy of a hydrogen atom at rest or in its lowest quantum state in the metal above its energy at rest, or in its lowest quantum state, outside.

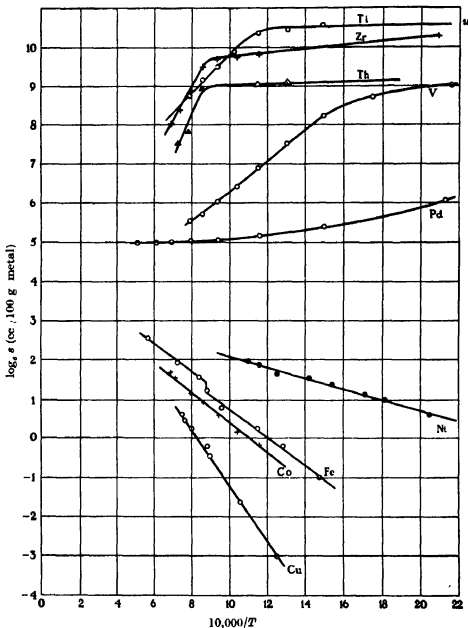


FIG. 1—Observed solubilities s of hydrogen in various metals subjected to one atmosphere pressure of H_2 , shown by plotting $\log_{10} s$ against $10^4/T$. The solubility s is the volume of H_2 gas (reckoned in c.c. at N.T.P.) absorbed by 100 g. of metal.

For sources. Cu (Röntgen and Möller 1934), Fe (Sieverts 1911); Ni (Smittenburg 1934); Co (Sieverts and Hagen 1934); V (Kirschfeld and Sieverts 1930); Ti (Kirschfeld and Sieverts 1929); Zr, Th (Sieverts and Roell 1926); Pd (Sieverts and Zapf 1935).

Let M_g be the number of hydrogen atoms and N_g the number of hydrogen molecules in volume V in the gas phase, and let $VF_1(T)$, $VF_2(T)$ be the partition functions for atoms and molecules respectively in this phase. Then we can write the equation of dissociative equilibrium in the form (Fowler 1936, p. 160)

$$\frac{M_g^2}{VN_g} = \frac{\{F_1(T)\}^2}{F_2(T)}. \quad (1)$$

The partition function $VF_1(T)$ has its standard form for a particle of mass m , and statistical weight ϖ_1 , namely (Fowler 1936, p. 57)

$$\frac{(2\pi mkT)^{3/2} V \varpi_1}{h^3}. \quad (2)$$

If the dissociation energy of the hydrogen molecule is χ_d , so that χ_d is the excess energy of two free hydrogen atoms at rest over their energy combined as a molecule at rest or in its lowest quantum state, then $VF_2(T)$ has the standard form (Fowler 1936, §§ 2.63, 3.4)

$$\frac{\{2\pi(2m)kT\}^{3/2} V \varpi_2 8\pi^2 AkT}{h^3 2h^2} e^{\chi_d/kT}, \quad (3)$$

in (3) ϖ_2 is the weight of the normal electronic state of the hydrogen molecule, A is its moment of inertia, vibrational states have been neglected and $8\pi^2 AkT/2h^2$ is the partition function factor for the rotational states, the factor 2 being the symmetry number. We may in these formulae consistently neglect (or include) nuclear spin weights. If we neglect them, then $\varpi_1 = 1$ and $\varpi_2 = 2$, this factor 2 arising from electron spin.

Let M_s be the number of hydrogen atoms in volume V of the metal, and let their partition function there be $VF_s(T)$. Then if classical calculations suffice, this has the form

$$\frac{(2\pi mkT)^{3/2} \varpi'_1}{h^3} e^{\chi_s/kT} \int_V e^{-W/kT} dV. \quad (4)$$

If nuclear spin factors are omitted ϖ'_1 is the electronic spin weight factor. The partition function (4) has the upper limit

$$\frac{(2\pi mkT)^{3/2} V \varpi'_1}{h^3} e^{-\chi_s/kT}, \quad (5)$$

and approaches this limit the closer, the higher T and the larger the regions in which W is small. It is proved in the appendix that (5) retains these limiting properties in the quantum theory.

3—We can now discuss the equilibrium balance of the hydrogen between the two phases. By the usual general theorems we have (Fowler 1936, p. 66)

$$M_g = \lambda VF_g(T), \quad M_o = \lambda VF_o(T), \quad (6)$$

and therefore

$$M_g = M_o F_g(T)/F_o(T). \quad (7)$$

The properties of the gas phase are determined by N_g rather than by M_g and we therefore use (1) to eliminate M_g , getting

$$\frac{M_g}{V} = \left(\frac{N_g}{V}\right)^{\frac{1}{2}} \frac{F_g(T)}{\{F_o(T)\}^{\frac{1}{2}}}. \quad (8)$$

If we write ν_s for the concentration of dissolved hydrogen atoms and p for the external hydrogen pressure, then

$$\nu_s \leq \left(\frac{p}{kT}\right)^{\frac{1}{2}} \frac{\frac{(2\pi mkT)^{\frac{1}{2}} \varpi'_1}{h^3} e^{-\chi_s/kT}}{\left[\frac{2\pi(2m)kT}{h^3}\right]^{\frac{1}{2}} \frac{8\pi^2 A kT}{2h^2}} e^{\frac{1}{2}\chi_d/kT}. \quad (9)$$

In (9) we have used the upper limit for $F_g(T)$ which might be fairly nearly approached in particular metals. The weight ϖ'_1 would be 2 if hydrogen atoms were really present as quasi-free structures in the metal. We shall see later, however, that this electronic spin factor 2 does not survive, and shall therefore put $\varpi'_1 = 1$, to avoid the need for later corrections to the calculations. Equation (9) then simplifies to

$$\nu_s \leq \frac{p^{\frac{1}{2}}}{(kT)^{\frac{1}{2}}} \frac{m^{\frac{1}{2}}}{2\pi^{\frac{1}{2}} h^{\frac{1}{2}} A^{\frac{1}{2}}} e^{-(\chi_s + \frac{1}{2}\chi_d)/kT}. \quad (10)$$

If we reckon the solubility s of hydrogen in the metal in the units of c.c. of molecular hydrogen at standard pressure and temperature (2.705×10^{19} molecules) dissolved per 100 g. of metal, then

$$s = \frac{\nu_s}{2 \times 2.705 \times 10^{19}} \times \frac{100}{\rho_m} = \frac{\nu_s}{5.41 \times 10^{17} \rho_m}, \quad (11)$$

where ρ_m is the density of the metal. Combining (10) and (11) and inserting the numerical values $m = 1.662 \times 10^{-24}$, $A = 4.63 \times 10^{-41}$ we find

$$s \leq [10^{1.91153}] \frac{p^{\frac{1}{2}}}{\rho_m T^{\frac{1}{2}}} e^{-(\chi_s + \frac{1}{2}\chi_d)/kT}. \quad (12)$$

For comparison with solubility measurements at one atmosphere ($p \sim 10^6$) and a mean temperature of 1000°K. in the metals of the iron group ($\rho_m \sim 8$) this reduces to

$$\log s \leq 5.89 - \frac{\chi_s + \frac{1}{2}\chi_d}{kT}. \quad (13)$$

The observations for the iron group metals conform to (13), those for Cu being exactly represented by (13) with the equality sign which is realized when the metallic partition function takes its upper limit.

4—Before proceeding it will be wise to verify, by considering the absorbed hydrogen as protons plus electrons, that we have in fact taken a partition function for the hydrogen in the metal of effectively the correct form. It is obvious that for the model we are using the most accurate assumption we can make is to take the dissolved hydrogen to be present in the form of dissolved protons, moving in a field of potential energy $\chi'_s + W$, and extra electrons which are added to the metallic electrons and distributed with these in the bands of metallic electron levels according to the Fermi-Dirac statistics. This is the simplest assumption that can be made. A more elaborate one was proposed and used by Herzfeld and Mayer (1934) in discussing solubility in Pd. To consider the phase equilibrium on this basis it will be sufficient to construct the various contributions to the free energy of Helmholtz's F , or to Planck's characteristic function Ψ ($= -F/T$). If N_0 is the normal number of metallic electrons in volume V to be considered, then the number in the hydrogen-rich metal is $N_0 + M_s$. The electrons make to the characteristic function of the metal the contribution (Fowler 1936, § 6.3)

$$k \left[Z - \mu \log \mu \frac{\partial Z}{\partial \mu} \right],$$

where

$$Z = \sum_r w_r \log(1 + \mu e^{-\epsilon_r/kT}),$$

summed over all the electronic states, of weight w_r and energy ϵ_r , and μ is determined by

$$N_0 + M_s = \mu \frac{\partial Z}{\partial \mu}. \quad (14)$$

Their contribution may therefore be written

$$\Psi_{e1} = k[Z - (N_0 + M_s) \log \mu],$$

which in virtue of (14) is such that $\partial \Psi_{e1} / \partial \mu = 0$. The protons make the simpler classical contribution

$$k M_s \left\{ \log \frac{V F'_s(T)}{M_s} + 1 \right\}.$$

There is also the contribution of the lattice vibrations, Ψ_{lat} , which we shall assume to be unaffected by the hydrogen. We find therefore

$$\Psi_s = \Psi_{lat} + k[Z - (N_0 + M_s) \log \mu] + k M_s \left\{ \log \frac{V F'_s(T)}{M_s} + 1 \right\}. \quad (15)$$

In writing $VF'_s(T)$ in (15) we have implicitly taken for the proton partition function its upper limit. The necessary correction can be made when required. In the same way hydrogen atoms contribute to the gas phase

$$\Psi_g = kM_g \left\{ \log \frac{VF_1(T)}{M_g} + 1 \right\}, \quad (16)$$

and equilibrium is given by the equation

$$\frac{\partial \Psi_s}{\partial M_s} = \frac{\partial \Psi_g}{\partial M_g}. \quad (17)$$

On carrying through the differentiation we see that (17) yields

$$\frac{M_s}{V} = \frac{M_g F'_s / \mu}{V F_1}, \quad (18)$$

in place of (7). Here F'_s is a proton partition function in place of the atomic partition function F_s , and μ is a parameter of the form $e^{\eta^*/kT}$, where η^* depends on the electron density and is the energy of the electrons at the top of the full levels in the Fermi-Dirac distribution at low temperatures. It will therefore strictly depend on M_s , but not significantly so long as $M_s/N_0 \ll 1$. Using (18) and (1) we find now

$$\frac{M_s}{V} = \left(\frac{N_g}{V} \right)^{\frac{1}{2}} \frac{F'_s(T)/\mu}{\{F_1(T)\}^{\frac{1}{2}}}. \quad (19)$$

The only essential difference between (19) and (8) is that χ_s now includes η^* and therefore may depend slightly on ν_s or s increasing as s increases, and that ω'_1 is necessarily 1, nuclear spin weights being neglected. We have already taken this value of ω'_1 , so that the preceding calculations are unaltered in form.

The energy χ_s which appears in (10) in ν_s can now be expressed explicitly in terms of other quantities. In virtue of (19) χ_s is replaced by $\chi'_s + \eta^*$, where χ'_s is the lowest energy of the absorbed proton and η^* the lowest possible energy of the absorbed electron. Together they represent the lowest energy of the absorbed hydrogen atom relative to a zero which refers to its lowest state as a hydrogen atom in the gas outside. By first ionizing the free hydrogen atom and then absorbing the proton and the electron we can express this energy excess as $I + \chi_p - \chi$, where I is the ionization energy of the H-atom (13.54 V), χ_p is the energy of solution of the proton and χ the thermionic work function of the metal. Thus in (10)

$$\chi_s = I + \chi_p - \chi. \quad (20)$$

This has been pointed out before by Frank (1933).

5—*Discussion of Observations for Ni, Fe, Co and Cu*—As we have already mentioned the observed solubility curves for these substances conform to (12) or (13). The observed curves, when plotted in the co-ordinates $\log s$ against $1/T$, are good straight lines whose slope determines $\chi_s + \frac{1}{2}\chi_d$. The observed values of $\chi_s + \frac{1}{2}\chi_d$ in cal./mole of hydrogen absorbed (heats of solution) are Cu, 14100, Co, 7300, Fe, 7000, Ni, 5600. Since χ_d is about 102,000 cal./mol., $\frac{1}{2}\chi_d$ is considerably greater than any one of these values and in all cases χ_s is negative, indicating that heat is evolved when hydrogen atoms are absorbed, as would be expected.

For Cu the observed values are such that one may say that they are accounted for by the theory if W/kT for protons inside the metal, at $T \sim 10^3$, is small compared with unity over regions comparable with the whole volume of the metal, while χ_s is about $-37,000$ cal./mol. For nickel χ_s is considerably larger numerically ($-45,400$) but applies to a much more restricted region of the metal, perhaps as little as $\frac{1}{4}$ of its total volume. From the fact that these solubility curves are good straight lines we may infer that the region of the interior of the metal in which $W \sim kT$ ($T \sim 10^3$) is small. The arrangement of the solubility curves is such that the numerically smaller values of χ_s (smaller heat of absorption per atom) hold good approximately over the whole interior of the metal, while the larger values hold only for more and more restricted regions. Clearly this correlation is what one should expect, as the occurrence of deeper and more restricted binding regions is the first move in the direction of the formation of hydrides.

6—*A simple theory for the case of hydride formation*—It seems probable that a deeper statistical discussion of hydrogen solubility when hydrides are formed will also prove fruitful and lead to a simple interpretation of the shape of the upper group of curves in fig. 1. An extremely tentative but simple version of such a discussion may be given as follows. As a first approximation we ignore the change of crystal form from the metal to the metal hydride lattice, and the fact that saturation does not in all cases correspond to the formation of the monohydride MeH , and assume that at all concentrations of hydrogen the solid can be treated as a perfect mixed crystal of M_h metal hydride molecules and $N_0 - M_h$ metal atoms. We shall assume further that the percentage of hydride is without effect on the lattice vibrations, but that the extra degrees of freedom due to the hydrogen in the lattice may be sufficiently accounted for by assuming a set of individual excited states for each proton relative to its own metal atom. This set of states leads to a partition function $v_s(T)$, let us say. On using the

standard formulae for a perfect mixed crystal we can then express Planck's characteristic function for the solid in the form (Fowler 1936, §§ 5.61, 6.3)

$$\Psi_s = \Psi_{\text{lat}} - kT M_s \log v_s(T) - kT \log \frac{N_0!}{M_s! (N_0 - M_s)!}. \quad (21)$$

The contribution of hydrogen atoms to the gas phase has the same value as before (equation (16)). On using Stirling's theorem in the last term of (21) and determining the equilibrium as in § 4 we find without difficulty that

$$\frac{M_s}{N_0 - M_s} = \frac{M_s v_s(T)}{V F_1(T)} = \left(\frac{N_0}{V} \right)^{\frac{1}{2}} \frac{v_s(T)}{\{F_2(T)\}^{\frac{1}{2}}} = \left(\frac{p}{kT} \right)^{\frac{1}{2}} \frac{v_s(T)}{\{F_2(T)\}^{\frac{1}{2}}}.$$

A reasonable expression for $v_s(T)$ is Planck's function for a three-dimensional isotropic oscillator of frequency ν multiplied by $w'_1 e^{-\chi_d/kT}$ for weight factor and allowance for the energy of the ground state. This gives

$$v_s(T) = w'_1 (1 - e^{-h\nu/kT})^{-3} e^{-\chi_d/kT}.$$

We have already seen that we must put $w'_1 = 1$ when we neglect nuclear weight factors. Thus

$$\frac{s}{s_0 - s} = \frac{M_s}{N_0 - M_s} = \left(\frac{p}{kT} \right)^{\frac{1}{2}} \frac{(1 - e^{-h\nu/kT})^{-3}}{\left[\frac{(2\pi(2m)kT)^{\frac{1}{2}} 8\pi^2 A kT}{h^3} \right]^{\frac{1}{2}}} e^{-(\chi_r + \frac{1}{2}\chi_d)/kT}. \quad (22)$$

We shall attempt a preliminary discussion assuming that it is accurate enough for the temperature range in question to take $1 - e^{-h\nu/kT} \simeq 1$. We shall actually find that quite reasonable variations in this factor from unity will enable us to account for the observations, but unfortunately we have no *a priori* means of estimating this ν , and we have of course omitted other effects which are doubtless of comparable importance. Inserting standard numerical values in (22) we find

$$\frac{s}{s_0 - s} = [10^{-1.22}] \frac{p^{\frac{1}{2}}}{T^{\frac{1}{2}}} e^{-(\chi_r + \frac{1}{2}\chi_d)/kT}. \quad (23)$$

For the standard conditions of one atmosphere pressure and temperatures near 700° K. we have

$$\frac{s}{s_0 - s} \simeq [10^{-3.2}] e^{-(\chi_r + \frac{1}{2}\chi_d)/kT}. \quad (24)$$

The omitted Planck factor if sensibly greater than unity will increase the numerical factor in (24).

7—(Added 10 March 1937) *Connexion between formulae (22) and (9).* It has been suggested to us by a referee that the connexion between these two methods of approximation, leading to formulae (22) and (9) should be

made clearer. Formula (22), and in particular its simplified version (23), is derived from an assumption of localized homes for protons in which they are tightly bound, formula (9) from an assumption of free movement of the proton throughout the volume of the metal. The connexion may be traced as follows

If $s \ll s_0$, which will certainly occur in a suitable temperature range if $\chi_s + \frac{1}{2}\chi_d > 0$, and if the vibrational partition function for the bound proton is not given a Planck form, but left unspecified as $v(T)$, formula (22) may be reduced to

$$s = s_0 \left(\frac{p}{kT} \right)^{\frac{1}{2}} \frac{v(T)}{\left[\frac{\{2\pi(2m)kT\}^{\frac{1}{2}}}{h^3} \frac{8\pi^2 A kT}{2h^2} \right]^{\frac{1}{2}}} e^{-(\chi_s + \frac{1}{2}\chi_d)/kT}.$$

If there are x homes for protons per metal atom, then

$$s_0 = \frac{100x}{m_{\text{Me}}} \frac{1}{5.41 \times 10^{17}}$$

in the former units of c.c. H_2 gas absorbed by 100 g. of metal, the gas being reckoned at normal pressure and temperature. In this formula m_{Me} is the mass of the metal atom. If v_a is the atomic volume, then $m_{\text{Me}} = v_a \rho_m$, and we can write

$$s_0 = \frac{1}{v_a/x} \frac{1}{5.41 \times 10^{17} \rho_m},$$

leading to

$$s = \frac{1}{5.41 \times 10^{17} \rho_m} \frac{v(T)}{v_a/x} \left(\frac{p}{kT} \right)^{\frac{1}{2}} \frac{e^{-(\chi_s + \frac{1}{2}\chi_d)/kT}}{\left[\frac{\{2\pi(2m)kT\}^{\frac{1}{2}}}{h^3} \frac{8\pi^2 A kT}{2h^2} \right]^{\frac{1}{2}}}. \quad (24')$$

On comparing (24') with (9) and (11) and remembering that $\varpi'_1 = 1$ we see that (24') is equivalent to (9) and (11) if and when

$$v(T) = \frac{(2\pi m k T)^{\frac{1}{2}}}{h^3} (v_a/x).$$

But this value of $v(T)$ is precisely that which one derives from the assumption that each proton is free to move in a field-free enclosure of volume v_a/x enclosed by infinite potential walls. The dimensions of the enclosure must not be less than 10^{-8} cm. in order that the spacing of the quantum states may be small compared with kT . The volumes v_a/x per proton home add up to the complete volume of the metal. It does not then matter, when the solubility is small, whether one assumes that the protons move freely throughout the metal or are each confined to their own proper fraction of the metal, the various fractions covering the whole volume—a well-known result. If the

regions in which $W \simeq 0$ are not coextensive with the whole metal one obtains for use with (24')

$$v(T) = \frac{(2\pi mkT)^{1/2}}{h^3} v' \quad (v' < v_a/x).$$

For larger values of s the two methods of approach naturally diverge, since one assumes that the proton excludes others from its own home, and the other continues to ignore all interactions. But the formula which we have applied to Ni for example can be equally well derived by either method of approach.

8 -*Discussion of observations for Ti, V, Zr and Th*—We shall naturally suppose for these metals that $-(\chi_s + \frac{1}{2}\chi_d) = \phi > 0$. Equation (24) then of course gives a solubility s which tends to the saturation value s_0 as $T \rightarrow 0$, which according to this simple version of the theory is equivalent to one hydrogen atom per metal atom. The actual saturated solubilities, shown in fig. 1, are approximately those corresponding to the compounds, TiH_2 , VH , ZrH_2 and ThH_3 respectively, but at this stage we make no attempt at discrimination.

A sharpish bend in the theoretical solubility should occur according to (24) when $s/(s_0 - s) \simeq 1$, or when

$$\phi/kT = 7.6. \quad (25)$$

We have to retain exact values of T in exponents. When $s \ll s_0$ we have

$$\log s = \text{const.} + \phi/kT,$$

so that the slope of the $\log s$, $1/T$ curves when they have steadied to straight lines on the high temperature side of the bend is also determined by ϕ/k . Finally as $1/T \rightarrow 0$ the solubility lines (if they actually continued straight) should cut the axis $1/T = 0$ at points such that

$$\log s = \log s_0 - 7.6,$$

that is in the same order and with the same spacing as they have when $T \rightarrow 0$.

A careful inspection of the curves of fig. 1 show that the observations do conform roughly to all these requirements of the theory. In fact, when we remember that the slopes of the curves for Zr and Th are not very well determined and that the omitted Planck factor may well be significant, the general agreement is suprisingly good and numerical estimates of ϕ in cal./mol may be made by both methods and give the results below.

	ϕ_{bend}	ϕ_{slope}
Ti	15,000	10,000
Zr	18,200	17,500
Th	18,600	22,500
V	9,100	7,700

The values of ϕ derived from the bend can be reduced by the Planck factor, but cannot be increased. The values here given for Th are therefore impossible on the theory, but the slope is obviously very poorly determined and might well be considerably smaller. If $(1 - e^{-h\nu/kT})^{-3} = e$ for the mean value of T concerned, the figure 7.6 in (25) is reduced by 1, so that quite reasonable values of ν could take up the remaining discrepancies.

It would be of considerable interest if a statistical calculation could be given covering the range of behaviour intermediate between the extreme cases here treated, and also including explicitly the formation of higher hydrides. On the experimental side it would be most illuminating if a comparative study of the solubilities of hydrogen and deuterium could be undertaken, as this would provide a number of important checks on the accuracy of the theory.

9—Appendix*—The particles are moving in an enclosure of volume V in which their potential energy is $W \geq 0$. We may suppose the potential energy to be introduced gradually, so that at any stage in its introduction we have the potential energy xW ($0 \leq x \leq 1$). In passing from x to $x + dx$ the change in the energy of any stationary state will be to the first order, according to standard theorems of perturbation theory,

$$\int_V \delta x W |\psi_x|^2 dV,$$

ψ_x being the wave function of the particle in that stationary state at that stage (x) of development of the potential energy. Since $W \geq 0$ this change is an increase. Therefore every eigenvalue of the energy increases steadily as x increases from 0 to 1. But the partition function $f(T)$ is $\sum_r w_r e^{-\epsilon_r/kT}$, and therefore as every ϵ_r increases $f(T)$ steadily decreases as x increases from 0 to 1. It follows that $f(T)$ takes its maximum value when $W = 0$.

REFERENCES

- Fowler, R. H. 1936 "Statistical Mechanics," 2nd ed. Cambridge.
 Frank, J. 1933 *Nachr. Ges. Wiss. Göttingen*, p. 293.
 Herzfeld, K. and Mayer, J. 1934 *Z. phys. Chem. B*, **26**, 203.
 Kirschfeld, L. and Sieverts, A. 1929 *Z. phys. Chem.*, **145**, 227.
 — — 1930 *Z. Elektrochem.* **36**, 123.
 Röntgen, P. and Moller, F. 1934 *Metallwirtschaft*, **13**, 81.
 Sieverts, A. 1911 *Z. phys. Chem.* **77**, 591.
 Sieverts, A. and Hagen, H. 1934 *Z. phys. Chem.* **169**, 237.
 Sieverts, A. and Roell, E. 1926 *Z. anorg. Chem.* **153**, 289.
 Sieverts, A. and Zapf, G. 1935 *Z. phys. Chem.* **174**, 359.
 Smittenburg, J. 1934 *Rec. trav. chim. Pays-Bas*, **53**, 1085.

* We have to thank Professor Dirac for this proof.

Complex Variables in Quantum Mechanics

BY P. A. M. DIRAC, F.R.S., *St John's College, Cambridge*

(Received 24 February 1937)

The general representation theory of quantum mechanics requires one to represent the states of a dynamical system by functions of a set of real variables q_r , each of which has a domain consisting of either discrete points or a continuous range of points, or possibly both together. A dynamical variable is represented by a function of two such sets of real variables q_r and q_s forming a generalized "matrix". In this paper we shall show that in certain cases it is advantageous to consider some of our variables q_r as complex variables and to suppose the representatives of states and dynamical variables to depend on them in accordance with the theory of functions of a complex variable.

In the usual theory the domains of the q_r 's are the eigenvalues of certain observables q_r . This significance of the q_r 's of course gets lost when we consider them as complex variables, but we have, however, some beautiful mathematical features appearing instead, and we gain a considerable amount of mathematical power for the working out of particular examples.

1—THE FUNDAMENTAL THEOREM

Suppose we have in our representation a variable q whose domain consists of all points from 0 to ∞ . This q may be, for example, the radius r in a system of polar co-ordinates, or a cartesian co-ordinate of a particle which is restricted to lie in a half of total space by an impenetrable plane barrier. The wave function representing any state will now be a function $(q|)$ of the variable q and of other variables which may be necessary to describe other degrees of freedom, which we do not need to mention explicitly.

Let us pass to a representation in terms of the momentum variable p conjugate to q . The wave function representing a state will now be

$$(p|) = \hbar^{-1} \int_0^{\infty} e^{-i p q / \hbar} dq (q|). \quad (1)$$

In the usual theory the variable p is restricted to be real, but we shall now allow it to be complex. We consider the function $(p|)$ as a function of a complex variable defined by (1) for all values of p in the lower half-plane

(i.e. all values of p for which the pure imaginary part of p is a negative multiple of i). We can easily see that *it must be regular in this domain*. For this purpose we note the physical requirement that the function $(q|)$ must be everywhere finite, or else contain singularities of a kind (such as the δ function) which give a finite integral when multiplied by a continuous function and integrated. Further, $|(q|)|$ must remain bounded as $q \rightarrow \infty$. It follows that when p is in the lower half-plane the integral (1) is absolutely convergent, and thus $(p|)$ cannot become infinite. Also, the function $(p|)$ defined by (1) in the lower half-plane is always single-valued and thus cannot have branch points in this domain. It follows that it must be regular in this domain.

By the principle of analytical continuation we may extend the domain of definition of our function $(p|)$ over the real axis into the upper half-plane. The formula (1) will still be valid on the real axis, except possibly at certain points where the function $(p|)$ has a singularity, but will not be valid beyond, unless $(q|)$ tends to zero very rapidly as q tends to infinity.

The conjugate imaginary of a wave function $(p|)$ is a function $(|p)$ which is regular for all values of p in the upper half-plane. A dynamical variable α is represented by a function $(p'|\alpha|p'')$ of the two variables p' and p'' , which is similar to $(p|)$ in its dependence on p' and to $(|p)$ in its dependence on p'' . Thus $(p'|\alpha|p'')$ is a regular function of p' in the lower half-plane and a regular function of p'' in the upper half-plane.

In our general scheme of quantum mechanics, whenever we have to do an integration over p , it will be of the type of an integral along the real axis

$$\int_{-\infty}^{\infty} (\dots|p) dp (p|\dots), \quad (2)$$

where the integrand contains two factors, the first $(\dots|p)$ being similar to $(|p)$ and therefore regular in the upper half-plane and the second $(p|\dots)$ being similar to $(p|)$ and therefore regular in the lower half-plane. From the theory of functions we know that we may distort the path of integration in any way provided we do not pass over a singularity in the integrand. We are thus led to our fundamental theorem, that *in performing an integration (2), we may choose any contour extending from $-\infty$ to ∞ such that the first factor of the integrand is regular everywhere above this contour and the second is regular everywhere below it*.

It may happen that the integral (2) along the real axis is indefinite as it stands, owing to the presence of certain kinds of singularity in the integrand on the real axis. Let us examine what interpretation quantum mechanics requires us to give to the integral in such a case.

The most important kind of singularity is a simple pole. Suppose we have a simple pole on the real axis in the second factor $(p| \dots)$. Then this factor must correspond to a wave function $(q| \dots)$ in the q variable of the form e^{iaq} for large q , a being some real number. To see this, we put

$$(q|) = \hbar^{\frac{1}{2}} e^{iaq}$$

in (1), obtaining

$$(p|) = \int_0^{\infty} e^{-i\mu q/\hbar} e^{iaq} dq$$

$$= \frac{i\hbar}{p - a\hbar} \left[e^{-i(\mu - a\hbar)q/\hbar} \right]_0^{\infty} \quad (3)$$

$$= \frac{-i\hbar}{p - a\hbar} \quad (4)$$

for p in the lower half-plane, and also for p on the real axis, except at the point $p = a\hbar$. (Thus involves neglecting the oscillating part of (3) for $q = \infty$, which is the usual practice in quantum mechanics.) Thus we have a simple pole at the point $a\hbar$.

When the second factor in (2) is of the form (4), to obtain the value of the integral (2) according to the usual procedure of quantum mechanics, we must first substitute for $(p|)$ the value (3) taken to some large definite upper limit g instead of to ∞ , and then, after evaluating the integral, make $g \rightarrow \infty$. This gives us for the critical part of the integral from $a\hbar - \epsilon$ to $a\hbar + \epsilon$, ϵ being small,

$$\int_{a\hbar - \epsilon}^{a\hbar + \epsilon} (\dots | p) dp (p|) = i\hbar \int_{a\hbar - \epsilon}^{a\hbar + \epsilon} (\dots | p) [e^{-i(p - a\hbar)g/\hbar} - 1] \frac{dp}{p - a\hbar}$$

$$= i\hbar \int_{-g\epsilon/\hbar}^{g\epsilon/\hbar} \left(\dots \left| a\hbar + \frac{\hbar x}{g} \right. \right) [e^{-ix} - 1] \frac{dx}{x},$$

where $p = a\hbar + \hbar/g \cdot x$. In the limit $g \rightarrow \infty$, this becomes

$$i\hbar (\dots | a\hbar) \int_{-\infty}^{\infty} [e^{-ix} - 1] dx/x = \pi\hbar (\dots | a\hbar).$$

This value for the integral (2) from the point $a\hbar - \epsilon$ to the point $a\hbar + \epsilon$ is the same as that which we should have obtained if we had done a contour integration along a small semicircle in the lower half-plane with centre $a\hbar$ and radius ϵ .

Thus we see that, when we are doing an integration of the type (2) in which there is a simple pole on the real axis in the second factor of the integrand, we must avoid the singularity by distorting our path of integration into the lower half-plane. Similarly, it may be shown that if there is a simple pole in the first factor of the integrand, we must avoid it by dis-

torting our path of integration into the upper half-plane. These distortions are in the directions allowed by our fundamental theorem.

Our theory would break down if both factors of the integrand had a simple pole at the same point on the real axis. This case, however, is never met with in practice. It would make the integral (2) infinitely great, according to the quantum-mechanical significance.

Poles of higher order than the first on the real axis will not ordinarily occur in our integrand (2), since a wave function $(p|)$ involving such a pole would correspond to a wave function $(q|)$ which increases without limit as $q \rightarrow \infty$ and is therefore not allowed physically. We may, however, sometimes wish to work with an operator whose representative $(p'|\alpha|p'')$ involves such a pole in p' or p'' . In these cases the operator will be defined so that, in performing an integration (2), we must avoid the singularity by distorting our path of integration from the real axis in the direction allowed by our fundamental theorem.

If a singularity of a different nature from a pole (i.e. a branch point without a pole superposed) occurs on the real axis, it will not give rise to any uncertainty in the integration through it. We may then distort the path of integration in the direction allowed by our fundamental theorem without altering the value of the integral. In this way we see that *our fundamental theorem is always valid, with any array of singularities on the real axis.*

2—CONDITIONS AT INFINITY

It is a rather ugly feature of our fundamental theorem that the path of integration must begin at $-\infty$ and end at ∞ . We might even sometimes meet with divergence at these points. It is thus necessary to make a closer investigation of the conditions at infinity.

Let us suppose first that both the factors in the integrand in (2) are regular at infinity, so that for large values of p they can be expanded in power series in p^{-1} , and let us suppose further that the constant terms in these series vanish. Then we have for large p

$$\begin{aligned} (p| \dots) &= a_1 p^{-1} + a_2 p^{-2} + a_3 p^{-3} + \dots \\ (\dots | p) &= b_1 p^{-1} + b_2 p^{-2} + b_3 p^{-3} + \dots \end{aligned} \quad (5)$$

Thus the total integrand is of the form

$$(\dots | p)(p| \dots) = c_2 p^{-2} + c_3 p^{-3} + \dots \quad (6)$$

If this is integrated over a large semicircle of radius R , extending from the point R to the point $-R$ in either the upper or the lower half-plane, the

result will tend to zero as $R \rightarrow \infty$. Thus we may take as our path of integration in (2) a closed contour, extending from $-R$ to R along the real axis and going back to $-R$ along a large semicircle in either the upper or the lower half-plane.

Let us now suppose one of the factors (5) contains a constant term, say the first, so that we have

$$(p | \dots) = a_0 + a_1 p^{-1} + a_2 p^{-2} + \dots \quad (7)$$

The integrand (6) will now contain the term $a_0 b_1 p^{-1}$ and (provided $b_1 \neq 0$) it will no longer have a precise meaning to integrate it from $-\infty$ to ∞ . Let us examine what meaning quantum mechanics would require us to give to this integral.

We must see what the important terms in (7) and the second of equations (5) correspond to in the q -representation. The term a_0 in (7) would be given, according to (1), by the q -wave function

$$(q | 0) = a_0 \hbar^{\frac{1}{2}} \delta(q),$$

the function $\delta(q)$ here being understood to lie entirely on the positive side of the point $q = 0$, so that

$$\int_0^{\infty} \delta(q) f(q) dq = f(0). \quad (8)$$

Again, the $b_1 p^{-1}$ term in the second of equations (5) would, according to the conjugate imaginary of equation (1), namely

$$(1 | p) = \hbar^{-\frac{1}{2}} \int_0^{\infty} (1 | q) dq e^{ipq/\hbar},$$

be given by the constant q -wave function

$$(1 | q) = -i \hbar^{\frac{1}{2}} \hbar^{-1} b_1.$$

The integrated product of these two wave functions is

$$\int_0^{\infty} (1 | q) dq (q | 0) = -2\pi i a_0 b_1 \int_0^{\infty} \delta(q) dq = -2\pi i a_0 b_1,$$

the δ function being interpreted in accordance with (8). From the transformation theory of quantum mechanics, the product must have the same value when evaluated in the p -representation, so that we must have

$$-2\pi i a_0 b_1 = \int (1 | p) dp (p | 0) = a_0 b_1 \int p^{-1} dp.$$

The domain of integration here is along the real axis, the pole at the origin being skirted, in accordance with our previous work, by a deviation into the upper half-plane, since the singularity occurs in the first factor of the integrand. It is now clear that, to get the right result, we must complete

our path of integration by a large semicircle in the lower half-plane and make it into a closed contour.

We can now see that, if a constant term appears in either of the expansions (5), we have to avoid the point at infinity in our path of integration in a somewhat analogous way to that in which we must avoid a simple pole on the real axis. If the constant term appears in the second factor, we must close up our path of integration by a large semicircle in the lower half-plane, and similarly if it appears in the first factor, we must close up our path of integration by a large semicircle in the upper half-plane. We cannot have a constant term in both factors. It would make the integral infinitely great, like a simple pole in both factors at the same point on the real axis.

We can now generalize our fundamental theorem to read as follows: *In performing an integration (2) we may choose any closed contour which divides the complex plane into two regions, such that the first factor of the integrand is regular in one of the regions (the one on the left-hand side of the contour), and the second factor is regular in the other. A constant term in either factor is here to be counted as having a singularity at the point at infinity.*

If either of the expansions (5) contains positive powers of p , these would correspond in the q -representation to terms involving derivatives of $\delta(q)$. Such terms do not seem to be of any practical importance as wave functions. We might possibly have to deal with an operator whose representative contains terms of this type, in which case we would define the operator in such a way that, when its representative occurs in an integrand, we must treat the singularity at infinity in accordance with our generalized fundamental theorem. Other independent types of singularity that may occur at infinity in the integrand would not affect the validity of our generalized fundamental theorem.

3—SOME SIMPLE OPERATORS

To obtain the representatives of operators in our complex p -representation, it is often not very convenient to make a direct transformation from the q -representation, but it is better to get the result by an argument which works entirely with the p -representation. The following work provides some simple illustrations of this.

We do not need to use the δ function in connexion with our complex variable p . We have the reciprocal function playing the part of the δ function. For example, the "unit matrix" is now

$$(p' | 1 | p'') = \frac{-i}{2\pi} \frac{1}{p' - p''} \quad (9)$$

To verify this, let us multiply this matrix into an arbitrary function $(p'|)$ in accordance with our fundamental theorem. The result is

$$\frac{-i}{2\pi} \int \frac{1}{p' - p''} dp'' (p''|), \quad (10)$$

the integral being taken along a closed contour such that the factor $1/(p' - p'')$ is regular in the domain on the left of the contour and the factor $(p''|)$ is regular in the domain on the right. Thus the simple pole at p' in the first factor must occur in the domain on the right of the contour, and must be the only singularity of the integrand in this domain. We may take the contour to be a small circle going clockwise round the pole at p' , and we then see that the integral has the value $(p'|)$. Thus the right-hand side of (9) plays the part of the unit matrix.

Similarly, the operator of multiplication by p is represented by the matrix

$$(p'|p|p'') = \frac{-i}{2\pi} \frac{p'}{p' - p''}, \quad (11)$$

since when this matrix is multiplied into an arbitrary function $(p'|)$ in accordance with our rules, the result will be (10) multiplied by the factor p' . It is a little surprising that the matrix (11) is not hermitian. Its conjugate, obtained by interchanging p' and p'' , and writing $-i$ for i , is

$$(p'| \bar{p} | p'') = \frac{-i}{2\pi} \frac{p''}{p' - p''}, \quad (12)$$

which differs from (11) by a constant. If we multiply the matrix (12) into an arbitrary function $(p'|)$, the constant will make its presence felt by the fact that the first factor in the integrand of

$$\int (p'| \bar{p} | p'') dp'' (p''|)$$

must be counted as having a singularity at infinity, and thus we must choose our path of integration such that the point at infinity is in the domain on the right, as well as the point p' . We may thus choose the path as in fig. 1, when the integral around the large circle will give the contribution of the constant part of (12).

To understand the physical significance of this constant part, we note that it is the representative of $i\hbar \delta(q)$, thus

$$\begin{aligned} i\hbar (p' | \delta(q) | p'') &= i\hbar \hbar^{-1} \int_0^\infty \int_0^\infty e^{-ip'q'/\hbar} dq' \delta(q') \delta(q' - q'') dq'' e^{ip''q''/\hbar} \\ &= i/2\pi \\ &= (p' | \bar{p} - p | p''), \end{aligned} \quad (13)$$

where $\delta(q)$ is assumed to lie in the domain 0 to ∞ , in accordance with (8). Thus the lack of hermitianness of the operator p is associated with a δ function at $q = 0$ and is of importance only when we operate on wave functions in q which do not vanish at the origin.

The dynamical variable q corresponds to the operator $i\hbar d/dp$. This is represented by the matrix

$$(p' | q | p'') = \frac{-\hbar}{2\pi} \frac{1}{(p' - p'')^2}, \quad (14)$$

because, when we multiply this matrix into an arbitrary $(p' |)$, we get

$$\frac{-\hbar}{2\pi} \int \frac{1}{(p' - p'')^2} dp'' (p'' |),$$

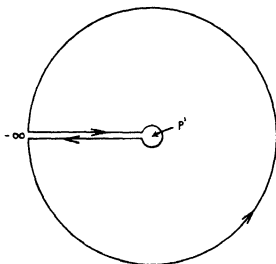


FIG. 1

the path of integration being the same as with (10), which gives

$$i\hbar \frac{d(p' |)}{dp'}.$$

Since q corresponds to $i\hbar$ times the operator of differentiation, q^{-1} must correspond to $-i\hbar^{-1}$ times the operator of integration. There is an arbitrary constant of integration associated with the operation of integration. This corresponds to the fact that we may add on an arbitrary multiple of $\delta(q)$ to the result of an operation of multiplication by q^{-1} . If we want to have no multiple of $\delta(q)$ in the result, we must choose our constant of integration

so that when our wave function is expanded in the form (7) the constant term vanishes. This is equivalent to putting the point at infinity as the lower limit in our integral.

We may represent q^{-1} by the matrix

$$(p' | q^{-1} | p'') = -\hbar^{-1} \log(p' - p'').$$

To verify this result, we note that when we multiply this matrix into an arbitrary $(p' |)$ we get

$$-\hbar^{-1} \int \log(p' - p'') dp''(p'' |), \quad (15)$$

where the path of integration must be such that the point p' and the point at infinity are in the domain on the right, and may thus be chosen as in fig. 1. The integrals round the small circle and the large circle will then vanish (from the condition that the function $(p' |)$ must not have any singularity in the domain on the right†) and we are left with the difference of the integral (15) from $-\infty$ to p' along two sheets of the log function, which difference is

$$-\hbar^{-1} \int_{-\infty}^{p'} 2\pi i dp''(p'' |) = -i\hbar^{-1} \int_{-\infty}^{p'} (p'' |) dp'',$$

as required.

4—APPLICATION TO THE HYDROGEN ATOM

The present theory enables us to use the powerful methods of the theory of functions of a complex variable in solving problems in quantum mechanics. These methods have already been used in many cases, notably in Schrödinger's original treatment of the hydrogen atom in 1926, but we have here put them on a systematic basis. As an example we shall now show how the treatment of the hydrogen atom appears in the present theory

We take the radius r to be our co-ordinate q , and we then have to solve the differential equation

$$\left[\hbar^2 \left(\frac{d^2}{dq^2} + \frac{2}{q} \frac{d}{dq} - \frac{n(n+1)}{q^2} \right) + 2m \left(W + \frac{e^2}{q} \right) \right] (q |) = 0, \quad (16)$$

† We require also the condition that the expansion of $(p' |)$ in the form (7) shall contain no p^{-1} term, as well as no constant term. If this condition is not fulfilled, the integral round the large circle will be infinitely great. The wave function $(p' |)$ then corresponds to a wave function $(q' |)$ which does not vanish at the origin and to which therefore the operator q^{-1} cannot legitimately be applied, as the square of the modulus of the resulting function would not be integrable.

W being the eigenvalue, and n being the order of the spherical harmonic concerned. Expressed in terms of the p variable, this equation reads

$$\left[-p^2 + 2 \int p + n(n+1) \int \int + 2mW - 2ime^2 \hbar^{-1} \int \right] (p|) = 0, \quad (17)$$

the integral sign being here used as an operator (in the sense of the preceding section with no constant term).

To obtain the solution of (17), we simply have to transcribe Schrödinger's solution of (16) into the p -representation. Schrödinger first eliminates the term $n(n+1)/q^2$ in (16) by making a transformation of the type

$$(q|) = q^\alpha (q|)^*$$

with

$$\alpha = n \text{ or } -n - 1.$$

This leads to

$$\left[\hbar^2 \left(\frac{d^2}{dq^2} + \frac{2(\alpha+1)}{q} \frac{d}{dq} \right) + 2m \left(W + \frac{e^2}{q} \right) \right] (q|)^* = 0, \quad (18)$$

which is Schrödinger's equation (7'). The corresponding transformation in the p -representation is

$$(p|) = \left(\frac{d}{dp} \right)^\alpha (p|)^* \quad \text{or} \quad \left(\int \right)^{-\alpha} (p|)^*, \quad (19)$$

according to whether α is positive or negative, and leads to the equation

$$\left[-p^2 + 2(\alpha+1) \int p + 2mW - 2ime^2 \hbar^{-1} \int \right] (p|)^* = 0. \quad (20)$$

The solution of (18) is given by Schrödinger's equation (12) in the form of an integral over a complex variable z , which is obviously playing the part of $1/\hbar$ times our variable p . The integrand here, apart from the factor $e^{i\mu z/\hbar}$, is therefore the solution of (20). Thus

$$(p|)^* = (p - c_1)^{\alpha_1 - 1} (p - c_2)^{\alpha_2 - 1}, \quad (21)$$

where c_1 and c_2 are the roots of

$$p^2 - 2mW = 0$$

and

$$\alpha_1 = -\frac{2ime^2/\hbar}{c_1 - c_2} + \alpha + 1, \quad \alpha_2 = -\frac{2ime^2/\hbar}{c_2 - c_1} + \alpha + 1$$

For large values of p , we have

$$(p|)^* = p^{\alpha_1 + \alpha_2 - 2} = p^{2\alpha},$$

and hence $(p|)$ is proportional to p^α . Now we require $(p|)$ not to have any singularity at ∞ and so we must take the negative value for α , namely $-n-1$.

For $n > 0$, we may use the other solution $\alpha = n$ provided we discard from it all the terms involving non-negative powers of p in its expansion in the form (5) in descending powers of p . One can easily see that this must give the previous solution again, since there can be no $p^{-1}, p^{-2} \dots p^{-n}$ terms in the $\alpha = n$ solution on account of its being, according to (19), the n -fold derivative of $(p|)^*$. The discard will be taken into account automatically if we arrange to have the point at infinity on the right-hand side of our contour of integration (instead of the left-hand side as it should be according to the fundamental theorem) whenever we use this solution in an integrand.

Our wave function $(p|)$ has singularities at the two points c_1 and c_2 . For any positive value of W , these points are both on the real axis and our wave function is alright. But for negative W , one of these points, say c_1 , will be in the lower half-plane, where our wave function is not allowed to have any singularity. We now get a permissible wave function only when the point c_1 becomes regular owing to α_1 being a positive integer, i.e. when

$$\frac{me^2/\hbar}{(-2mW)^{1/2}} = n$$

is a positive integer, s say. This leads to the Bohr formula

$$W = -\frac{me^4}{2\hbar^2(n+s)^2}.$$

The case of $n = 0$ is a little exceptional in that then (21) is not really a solution of (20). This may be seen from the fact that (21) is now of the form p^{-2} for large p , so that the first term of (20) will contribute a constant for large p , and no other term in (20) can contribute another constant to cancel with this one. Thus, we should have a constant term on the right-hand side of (20), which would correspond to a $\delta(q)$ term on the right-hand side of (18), implying a failure of our solution at the origin. The fact that this solution is allowed by quantum mechanics shows that our theory does not always automatically give the correct boundary conditions at the origin.

The above example shows the great superiority of the p -representation in dealing with this kind of problem, in that it allows the wave function $(p|)^*$ to be expressed in finite form. In fact, we may say that Schrödinger's (1926) treatment is effectively a treatment in the p -representation, his complex variable z playing the part of our p . Any subsequent calculations that we may wish to make, such as the evaluation of matrix elements, may be conveniently carried out with the wave functions $(p|)^*$ and the machinery

of contour integration. The work will be specially simple for the discrete states, since the wave function then has only one singularity, and any contour integration may be performed round a small circle enclosing this singularity.

REFERENCE

Schrodinger, E. 1926 *Ann. Phys., Lpz.*, **79**, 361-76.

On the Pattern of Proteins

BY D. M. WRINCH, M.A., D SC., *Mathematical Institute, Oxford*

(Communicated by R. Robinson, F.R.S.—Received
8 July 1936—Revised 19 January 1937)

1—INTRODUCTION

Any theory as to the structure of the molecule of simple native protein must take account of a number of facts belonging to many different domains of science. As our starting point we take the following:

(a) The molecules are largely, if not entirely, made up of amino- and imino-acid molecules. They contain peptide linkages, but in general few —NH_2 groups not belonging to side chains, and, in some cases, possibly none (Cohn 1928).

(b) There is a general uniformity among proteins of widely different chemical composition (Jordan Lloyd 1926; Astbury and Lomax 1935); presumably, therefore, there is a simple general plan in the arrangement of the amino- and imino-acid residues characteristic of proteins in general.

(c) A large number of crystalline proteins have highly symmetrical crystals (Schimper 1881); unlike the low symmetrical forms of most organic compounds, these usually possess triad or hexad axes. Insulin (Crowfoot 1935) and pepsin (Bernal and Crowfoot 1934) crystals are of trigonal type. In the case of insulin this symmetry has been shown to be possessed by the molecules themselves (Crowfoot 1935). In other cases also, by analogy

with other trigonal crystals, the molecules may be expected to have pseudo-trigonal symmetry.

(d) Many proteins are "globular" in form (Svedberg 1930a, 1930b, 1933, 1934).

The facts (a), (b), (c) suggest that the molecule of native protein may contain closed polypeptides (i.e. polypeptides with no open ends); that the polypeptides, open or closed, fall into rings; and that the ring system is such as to allow the possibility of a regular and orderly arrangement of hundreds of residues which, to some extent at least, is independent of the particular amino and imino acids in the molecule.

These suggestions can be given precise formulation in more than one way. Two in particular seem worthy of detailed consideration. The formulation in terms of polypeptides folded by means of hydrogen bonds, which has already been worked out, forms the subject of another communication (Wrinch and Jordan Lloyd 1936). In this communication we formulate the suggestions in terms of what may be called "cyclized" polypeptides (Wrinch 1936a).

An investigation of the properties of such molecules shows that, geometrically, they may equally well be regarded as polymers of substituted diketopiperazine molecules. It is considered to be an important feature of the present hypothesis that the conflict between the "polypeptide" and the "diketopiperazine" schools of thought in protein chemistry is automatically resolved. Both views find a common fundamental expression in terms of a *laminar molecule*, an open or closed surface polymer of amino- and imino-acid residues.

We call this surface polymer a laminar molecule, whether it be open or closed, in order to indicate that it is two dimensional, in the sense that it is but one residue thick. The closed lamina at once suggests a globular type of molecule. The theory devised to cover the facts (a), (b) and (c) thus automatically covers the facts (d).

As a prelude to an investigation of this and all the possible types of lamina which can be built with these residues, this communication concentrates attention on those which lie approximately in one plane. The approximately plane laminar molecule certainly appears to offer an interpretation of protein films one residue thick, and in particular of the fact that different proteins have films with characteristically different densities which cannot be accounted for on the ground of a difference in the average residue weight. Furthermore, the residues which make up the lamina possess characteristic side chains (or rings), and the lamina itself possesses hydroxyls. Lamina may therefore be linked to lamina by side chains

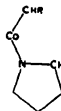
(or rings) and by hydroxyls, and a laminate aggregate several residues thick can be built.

The geometrical aspect of the situation shows at once, unequivocally, that in the process of surface polymerization of amino- and imino-acid residues it is not necessary to introduce into the protein molecule any essentially new type of link. It is sufficient simply to postulate the existence of multiple peptide links, a generalization of Emil Fischer's peptide link which 30 years of intensive research in organic chemistry has confirmed as the fundamental link in the protein molecule. Researches belonging to the domain of physical chemistry, however, disclosed the existence of "globular" proteins (Svedberg 1930*a*, 1930*b*, 1933, 1934). The peptide link has been viewed as a method of building linear polycondensations of amino and imino acids. How then, can this link be the predominating link in the "globular" protein? A satisfactory answer to this difficulty is supplied by the simple geometrical deductions which follow from the idea of the closed laminar molecule.

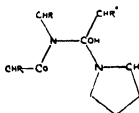
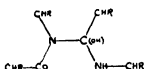
It must be recognized that on the present theory proteins have a ring structure (though some of the constituent cyclized polypeptides may be open, i.e. have open ends). This idea, which has often been suggested (Fischer 1906, Abderhalden 1923, Abderhalden and Komm 1924*a*, 1924*b*, 1924*c*; Bergmann and others 1924, 1925, 1926*a*, 1926*b*) and appears *a priori* to be quite acceptable, has apparently been discounted from the start, possibly owing to a failure to realize that whereas amino and imino acids readily form linear polycondensations of considerable length by means of the peptide link, there is also no difficulty in their forming surface polycondensations of considerable extent by a similar mechanism. Theoretically, this is certainly the case. Line polycondensation of these units proceeds by means of classical peptide links. These are *single peptide links*, in which a single pair of C and N atoms are joined to form a new atomic complex. Surface polycondensations should then proceed by means of *multiple peptide links*, the double peptide links, and the triple peptide link, in which two pairs or three pairs of C and N atoms are joined to form new atomic complexes.

In order to develop the theory to the point at which it may be tested by experiment, not only in enzyme chemistry, but also in organic, physical and immuno-chemistry, it is necessary to investigate the structural properties of single laminae and laminate molecules and molecular aggregates. The general plan has been to devote §§ 2-5 of this paper to the approximately plane laminar molecule, which finds its chief application in the field of protein films. Some of the applications of the theory to protein molecules

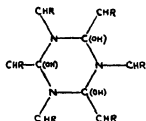
in general, which depend essentially upon the laminae of which they are composed, are also considered in these sections. Those problems whose treatment depends essentially upon the ways in which these laminae are combined are then discussed in §§ 6-10, in which laminate structures in general are considered.



1 1 single peptide links



1 2 double peptide links



1 3 triple peptide link.

Finally, it should be made clear that we do not claim that the present treatment of the problem is in any sense complete—the paucity of data in any case makes this impossible at the present time. Nor do we claim that the solution offered is, in the mathematical sense, unique. An alternative (or supplementary) treatment by means of the postulate of hydrogen bonds (Wrinch and Jordan Lloyd 1936) has already been given. At this early stage of the work, we confine attention to molecules and molecular aggregates built from certain approximately plane surface polymers of the amino- and imino-acid residues, which certainly satisfy *some* of the requirements of protein chemistry. In the happy event of this line of approach to the protein problem proving fruitful, the polymer postulate in all its forms will be given a full and exhaustive treatment.

2—THE CYCLIZED POLYPEPTIDES

An examination of the geometrical nature of polypeptide chains shows that they may be folded into hexagonal arrays. We begin with closed polypeptides, i.e. polypeptides with no open ends. A chain containing two residues, the substituted diketopiperazine molecule, forms a single hexagon. A chain consisting of six residues falls naturally, geometrically speaking, into the neat hexagonal structure shown in fig. 1. Certain longer closed chains, for example those containing 18 or 42 residues, also fall into structures of the same general hexagonal type. On this occasion we work out the implications of this type of folding, leaving aside a second type of folding based upon hydrogen bonds between C'-O and HN groups

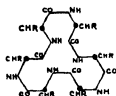


FIG 1

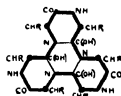
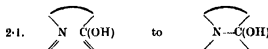
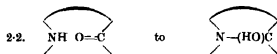


FIG 2. A "cyclol 6" molecule.

The stability of these folded polypeptide chains (as pictured in fig. 1) cannot be attributed to electrostatic attractions between the various CO and NH groups, since the appropriate distance between C and N atoms in these circumstances lies between 2.8 Å and 4.2 Å,* whereas the distance in our case is at most 1.54 Å. If, however, the C'-O-NH groups are in the lactim form C(OH)-N and we postulate an intramolecular "linkage" transformation



or if we postulate an intramolecular (prototropic) transformation†



the situation is, atomically speaking, "regularized" and we obtain (fig. 2)

* International Tables for the Determination of Crystal Structure (1934).

† The application of this transformation to these molecules was suggested to me by Dr J. D. Bernal. I understand from Dr W. T. Astbury that the transformation itself was first suggested by F. C. Frank in 1933, in connexion with keratin. See Astbury (1936).

the tetracyclic molecule "cyclol 6" by the "cyclization" of the closed polypeptide consisting of 6 residues.

In this communication, no decision is reached as to which of the suggested mechanisms is more plausible. Whichever is adopted—and in a sense the distinction is more apparent than real—the transformations reduce the number of double bonds per residue in the molecule and restore to the carbon atoms their natural tetrahedral complement of surrounding atoms.

It is desirable that the energy balance of the process of formation of cyclol molecules from polypeptides should be fully investigated in due course, since these considerations afford one method of assessing the probability of the present theory. The precise data which are required are not yet available, but the case of cyclol 6 has already been discussed in a preliminary manner (Wrinch 1936*b*), in the hope of calling attention to the fact that certain data are urgently required.

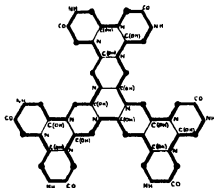


FIG. 3—A "cyclol 18" molecule.

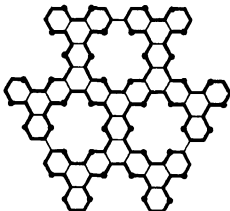


FIG. 4—A "cyclol 42" molecule

The molecules cyclol 6, cyclol 18, cyclol 42, shown in figs 2, 3, 4, are some of the early members of a series of cyclol molecules containing 6, 18, 30, 42, 54, 66, 78, 90, 102, ... $(6 + 12n)$, .. residues respectively. If we leave the side chains out of account, these molecules possess threefold central symmetry. There is also a companion series with twofold central symmetry, consisting of 10, 18, 26, 34, 42, 50, 58, 66, 74, 82, 90, 98, ... $(2 + 8n)$, ... residues respectively. Series with sixfold central symmetry and with no central symmetry can also be constructed.

Open polypeptides may also form polycyclic structures by means of cyclization. Thus the nonapeptide, shown in fig. 5, has formed a single hexagon by means of one intramolecular linkage transformation of this type; it is partially cyclized. By means of four such transformations it

forms a structure containing 4 hexagons, as shown in fig. 6: here it is completely cyclized. Open polypeptides, no matter how long they may be, can form polycyclic structures after the same manner. Such polycyclic structures can have open ends. Throughout this communication the assumption of a ring structure for proteins is not to be interpreted to mean that there are no open ends.

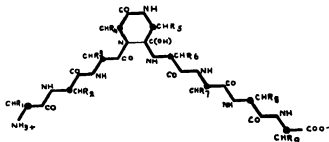


FIG. 5

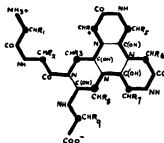
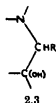


FIG. 6

It is a necessary condition for the cyclization of polypeptides into hexagonal arrays, by means of one or other of the suggested intramolecular linkage transformations, that the amino and imino acids concerned are all α -derivatives. The hypothesis that the polypeptides in proteins are cyclized therefore implies that all the amino and imino acids in proteins are α -derivatives. This is known to be the case. With the exception of β -alanine, a constituent of carnosine, and of its derivative anserine, all the amino and imino acids occurring in nature are of α -type, these peptides β -alanyl histidine and β -alanyl methyl histidine have been found in the muscles of all the vertebrates which have so far been studied, they have never been derived from proteins. A remarkable uniformity in nature thus finds an unforced interpretation and significance in terms of the present theory of protein structure.

The folding of polypeptides into hexagonal arrays which is now under consideration suggests a new building unit for the protein molecule



which may evidently be used to construct molecules of a variety of types. These are being investigated in detail. In this preliminary communication we make the assumption that the valency angles for N atoms and for C atoms are the tetrahedral angle. According to X-ray determinations, we may take the distances C—N, C—C as 1.42 Å (Wyckoff and Corey 1934), and 1.54 Å respectively. We therefore, for the moment, use a mean value $a = 1.48 \pm 0.06$ Å for these three distances indifferently. Both the diazine and the triazine hexagons (see fig. 2) can then be approximately plane, of cyclohexane type. We call the plane through the mid-points between consecutive atoms of a hexagon the "median" plane of the hexagon. One possibility is then that each trio of diazine hexagons is built into a central triazine hexagon (e.g. in cyclol 42) in such a way that all hexagons share the same median plane, and the whole molecule, no matter how far we go in the series of cyclols, is itself approximately plane. For many purposes it is adequate to replace the atoms in the hexagons by their projections in the median plane. We then have to do with arrays of hexagons in which the side of every hexagon is $b = 2\sqrt{2}a/3 = 1.395 \pm 0.055$ Å, so that the corresponding diameter is 2.79 ± 0.11 Å.

On this occasion we confine attention to these "puckered", nearly plane laminar molecules, reserving for a future publication an investigation of the molecules which are not even approximately plane. In any case a full discussion of cyclol structures must wait upon a detailed X-ray analysis of the appropriate compounds. Thus an analysis of various substituted diketopiperazine molecules, which in the particular case of glycine anhydride have been investigated and shown to be "nearly flat" (Bernal 1931), would guide us in the geometry of the diazine hexagons in the cyclol molecule. For the triazine hexagons, an investigation of the derivatives of cyanurol might be illuminating.

The arrangement in space of the side chains belonging to these laminar molecules depends upon the stereochemistry of the naturally occurring amino acids derived from proteins (Jordan Lloyd 1932). This subject has

received much attention since Fischer and Raske (1907, 1908) made the first contributions to it in 1906. The conclusion reached by a number of workers including Karrer and Kaase (1919, 1920), Clough (1918) and Wasser and Brauchli (1924) is that all amino and imino acids obtained from proteins have the same configuration. If we use the usual conventions (according to which the fifth carbon atom of *d*-glucose has the *d*-configuration) all the amino and imino acids derived from proteins are of laevo type as shown in fig. 7. In this figure the atoms N—C—C lie in the plane of the paper; the direction CO to NH is clockwise and the side chain *R* emerges above the paper. Throughout this paper a residue of laevo



type will be conventionally represented by $\text{N} \diagup \text{C}$

This is the second remarkable uniformity which has been found in all the amino and imino acids derived from proteins. It also has considerable significance on the cyclol hypothesis. This uniformity of configuration implies that the cyclol molecule is dorso-ventral, i.e. it has two surfaces, the "front" surface from which side chains emerge, and a "back" surface free from side chains. The hypothesis therefore explains the stability on a water air interface of protein films one residue thick, and so provides the logical link between two sets of experimental results.

3.—THE CLOSED 'CYCLOLS'

The fact that the number of free NH_2 groups not belonging to side chains in the protein molecule is small, even in some cases zero, confers on the closed polypeptides some measure of importance. Furthermore, many of the properties of cyclized polypeptides in general can be studied more easily in the particular case of closed polypeptides. We confine the term "closed polypeptide" to a single chain, and the term "cyclol" molecule to those polycyclic structures which, whatever their type of symmetry, are each made up of a single closed completely cyclized polypeptide.

We begin with, and on this occasion we limit ourselves to, the cyclols which, apart from the side chains of the constituent residues, have three-fold central symmetry. All the residues in this case, as throughout this paper, are taken to be of laevo type. Cyclol 6, the first member, consists of three diazine hexagons, built into one central triazine hexagon. There are two isomeric forms, one of which is shown in fig. 8. In both cases the side chains emerge above

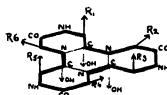


FIG. 8

the median plane of the molecule. In one isomer (see fig. 8), side chains R_1, R_2, R_3 , one belonging to each diazine hexagon, emerge normal to this plane, the others, R_4, R_5, R_6 , at a small angle to it, and vice versa in the other isomer. The three hydroxyls emerge normal to the plane, in one isomer below, as in fig. 8, in the other above.

The characteristics of cyclol 18 are similar. In this case there are now four triazine rings. In one isomer the hydroxyls of the central triazine ring emerge above the paper, those of the other three triazine rings emerge below the paper. In the other isomer the directions of the hydroxyls are reversed. In the case of all the cyclol molecules the hydroxyl trios associated with the triazine hexagons are arranged after the same manner, so that a trio above the paper is surrounded by three trios below the paper and vice versa no matter how many such trios there may be. Thus, in cyclol 42, shown in fig. 4, the central trio is surrounded by three trios with the opposite sense, each of these is surrounded by three trios with the original sense.

In the case of cyclol 42 (fig. 4) we see for the first time in this series the building of hexagonal lacunae by cyclized polypeptides which contain a sufficient number of residues. In this molecule three such lacunae are arranged symmetrically about the centre of the molecule. Each lacuna is fringed by six side chains, all emerging above the median plane of the molecule, three normal to the median plane of the molecule, and three at a small angle to it. The existence of these hexagonal lacunae of standard size and the arrangement of these lacunae in hexagons is a characteristic of the pattern of cyclized polypeptides in general.

The fact that one trio of the side chains, fringing such a lacuna, emerges above the median plane, making only a small angle with the median plane of the molecule, can later be used as a guide in the arrangement of individual residues.

The question of the imino acids—proline and hydroxyproline—is also relevant at this point. In these cases the residues take the special imino form instead of the standard form for amino acids (see 1.1): an imino-acid residue can therefore play no part in forming a triazine ring (see 1.2 and 1.3). The proportion of imino acids in a protein is likely to be of value in working out its structure. It is never so high as to rule out the possibility of some triple peptide links.*

We have here given indications of the way in which the general cyclol pattern offers guidance as to the arrangement of this and that type of

* Bergmann (1935) finds the proportions of proline and hydroxyproline in gelatin to be one-sixth and one-ninth respectively.

residue. A detailed discussion of the matter will be given on another occasion. At the moment we would stress the fact that the various cyclol molecules containing an ever-increasing number of residues represent but one single general pattern. This is shown in fig. 9. The characteristics of this pattern may be summed up as follows. It consists of diazine and triazine hexagons, built into one another in such a way that every triazine

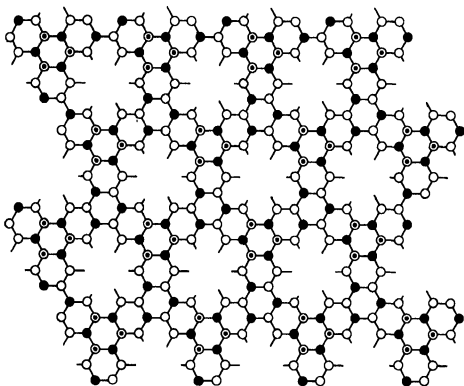


FIG. 9—The cyclol pattern. ● = N. ○ = C(OH), hydroxyl upwards. ⊙ = C(OH), hydroxyl downwards. ○— = CHR, direction of side chain initially outwards. ⊙— = CHR, direction of side chain initially upwards. The median plane of the lamina is the plane of the paper. The lamina has its "front" surface above and its "back" surface below the paper.

hexagon is surrounded by three of the diazine type and pairs of triazine hexagons are joined by a single diazine hexagon. In a sense we may think of the triazine hexagons as units of order 3, the diazine hexagons as units of order 2. The pattern depends upon the fact that every unit of order 3 is surrounded by three units of order 2, while every unit of order 2 joins two units of order 3. The formation of characteristic lacunae is then logically entailed. The other characteristics of the pattern are concerned with the

arrangement of the hydroxyl groups in alternating hexagonal arrays, and of the side chains in such a way that the "back" surface of the molecule is entirely free from them.

4—POLYMERIZATION AND CYCLIZATION *

The geometrical characteristics of cyclol molecules suggest that they lend themselves readily to polymerization *

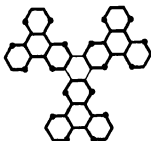


FIG. 10 A "cyclol 6" trimer.

In fig. 10 there are three cyclol 6 molecules having the same configuration. By means of new links between C and N atoms they can trimerize into a molecule with 18 residues, which is simply cyclol 18. Presumably cyclol 6 molecules can polymerize indefinitely as in fig. 11 and so form a laminar

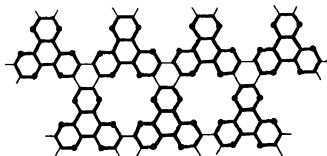


FIG. 11 —A "cyclol 6" polymer.

molecule with the characteristic cyclol attributes, covering an area of any shape and extent. Polymerization of this type will occur only among cyclol molecules having the same configuration.

In a similar way higher cyclol molecules may be presumed to polymerize

* Here as elsewhere the argument is concerned only with what is geometrically possible.

readily. A polymer of cyclol 18 is shown in fig. 12. Polymers of cyclol 42, cyclol 66, ... can also be constructed. Mixed polymers consisting of cyclol 6 and cyclol 18 molecules (for example) in certain proportions can also be built.

An interesting point about these polymers resides in the fact that each cyclol molecule can build only certain polymers, each characterized by certain types of lacunae. A characteristic density is correspondingly implied. Thus in the polymer of cyclol 6 shown in fig. 11 the area per residue, with $b = 1.395$, is $B = 3\sqrt{3}b^2 =$ (say) 10 \AA^2 . This evidently represents

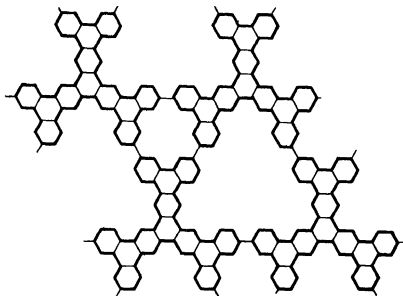


FIG. 12—A "cyclol 18" polymer.

the highest density for any cyclol molecule or its polymers. In the polymer of cyclol 18 shown in fig. 11 the area per residue is $4\sqrt{3}b^2 = 4B/3$. After the same manner cyclol 66 can build a polymer in which the area per residue is $18B/11$.

The consideration of the polymers formed by the cyclol molecule and of the cyclol pattern in general (fig. 9) makes it perfectly clear that a similar structure could be obtained by the polymerization of diketopiperazine molecules as shown in fig. 13. The present theory does not imply that the laminar molecules are in fact formed in this way. The implications of the theory with respect to the degradation of proteins are under consideration. It is already plain that the probability of obtaining diketopiperazine molecules from the laminae is small.

I have already given a discussion of the energy balance of the process of formation of cyclol 6 from the closed polypeptide containing six residues (Wrinch 1936*b*). This is immediately applicable to the trimerization of cyclol 6 and can be extended to cover the cyclization and polymerization processes in general. As a preliminary to such an investigation, we notice that a useful index is supplied by the geometry of the molecules themselves, namely the number of (CO, NH) groups divided by the number of residues in the molecule.

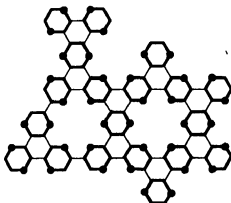


FIG. 13—A dikotopiperazine polymer.

Thus cyclol 6 has one per two residues; the index is $\frac{1}{2}$. Cyclol 18 reduces this index to $\frac{1}{3}$. Figures for higher cyclols are given in Table I.

TABLE I

Number of residues	Number of (CO, NH) groups	Index
6	3	0.5
18	6	0.33
30	7.5	0.25
42	9	0.214
54	12	0.222
66	12	0.182
78	15	0.192
90	16.5	0.183
102	19.5	0.191
114	21	0.184
126	22.5	0.179
138	24	0.174
150	27	0.180
162	27	0.167
174	30	0.171

The significant point will be noticed that this index is not a monotonic decreasing function when a series of cyclols with ever-increasing numbers of residues is considered. Picking out for each number of residues the cyclol for which the index is lowest, we notice that, at certain points in the sequence of molecules, the index increases.

In a similar manner, the formation of polymers of cyclol molecules in general reduces the number of (CO, NH) groups per residue in the resulting polymer. Cyclol 6 has $\frac{1}{2}$; cyclol 18, the trimer of cyclol 6, has $\frac{1}{3}$. The polymer of cyclol 18, shown in fig. 12, reduces the index still further, indeed, the index when the polymer is indefinitely continued is $\frac{1}{12}$. Similarly the polymer of cyclol 6 shown in fig. 11, when indefinitely continued, reduces the index to zero. The situation may be summed up roughly by saying that polymerization and cyclization replace a surface distribution of (CO, NH) groups by a line distribution.

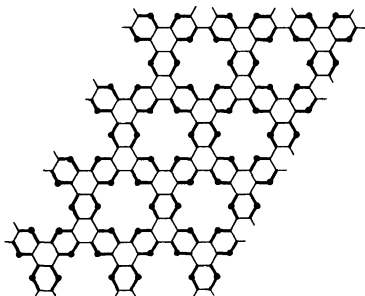


FIG. 14—Surface polycondensation of amino acids.

These preliminary calculations tend to indicate that we may have to consider not simply cyclol molecules themselves, but polymers of cyclols in general, including mixed polymers.

One further simplification of the whole situation now presents itself in that cyclol polymers, as also cyclized polypeptides, are themselves, formally speaking, surface polymers of amino- and imino-acid residues. The densest surface polymer—the “cyclol” pattern—is shown in fig. 14. Other surface

polymers are simply this pattern with certain proportions of the residues missing. No matter which of the various units be taken as our starting point in protein synthesis, the essential topic is the formation of polymers of the residues. It follows that a complete investigation will consider not only the cyclized polypeptides and their polymers, but all the molecules which can be built by surface polymerization, and if necessary volume polymerization of the residues

The linear polymers of these residues are closed (uncyclized) polypeptides, which possess one (CO, NH) group per residue. Their index is therefore one. The surface polycondensations on the other hand, comprising the cyclols and the cyclized polypeptides and their polymers, have lower indices. Indeed, by a suitable arrangement of the residues, the index can be greatly reduced. A general investigation of all the possible open and closed surface polymers of the residues is now being undertaken. It already indicates that in the set which have, for example, a threefold symmetry axis, there exist certain surface polymers with specially low indices, these consist of certain definite numbers of residues. It also calls attention to the fact that imino-acid residues, being unable to take part in the formation of triazine hexagons, help to bring cyclization and polymerization to a natural conclusion.

We view the surface polymers of the residues as the fundamental entity in the protein molecule, a few such entities being combined together in certain cases in two ways which will be discussed later. On the basis of these considerations with respect to (1) the indices of such polymers, and (2) the proportion of imino residues, we deduce that proteins in general fall into a few molecular weight classes (when allowance is made for the different average residue weights in different proteins). Similar deductions hold also for closed surface polymers.

The steps in this deduction run as follows. Surface polymers of the amino- and imino-acid residues, like the line polymers, are in general unsaturated molecules. Only those consisting of certain definite numbers of residues and certain definite proportions of imino-acid residues tend to exist in isolation. The essential entities in the protein molecule, the surface polymers, therefore fall into certain molecular weight classes (when allowance is made for the different average residue weights). The proteins themselves, made up of one or more such entities, therefore similarly fall into certain molecular weight classes. [*Note added in proof 31 March 1937. These deductions are explained in a communication shortly to appear in Nature.*]

These deductions from the present theory are confirmed by a group of

facts established by Svedberg. We quote Svedberg's own words (Svedberg 1933). "We have found that the molecules of most of the homogeneous native proteins are simple multiples or submultiples of 34,500, which is the molecular weight of ovalbumin. Only a very limited number of different molecular weights are represented among the proteins. On the other hand, we know a large number of proteins differing widely with regard to chemical composition, isoelectric point and light absorption. This means that chemically different proteins may have the same (or nearly the same) molecular weight. As a matter of fact we find that the numerous proteins fit into a few molecular weight classes. Recent investigations have shown that this regularity probably obtains from the lowest molecular weight so far observed for a protein, or 17,000, up to the highest weight, or 5,000,000. It seems that just about a dozen different steps are required in order to proceed from the lowest to the highest weight. With increasing weight the absolute interval between the steps becomes larger and larger. Thus there are six different molecular weights between 17,000 and 200,000 and also six between 300,000 and 5,000,000. What reality chemical or physiological is the explanation of this puzzling display of figures we do not know as yet."

The present theory suggests that the explanation is to be found—not in chemical or physiological considerations—but in the abstract geometry of polymers of amino- and imino-acid residues.

5—PROTEIN FILMS

So far we have been concerned to build polypeptides into single laminae. Our next task is to consider how these laminae, which may presumably be of considerable extent, can be built into a laminate molecule or molecular aggregate by the superposition of lamina upon lamina. Before we do this it is natural to consider what knowledge is available as to the existence of proteins in the form of a single lamina. This brings us into the field of protein films. A considerable amount of information with regard to these films is now available. Unimolecular films of gliadin, glutenin, egg albumin, zein, serum albumin and serum globulin have been obtained, the densities of which range from 1.724×10^{-7} g./cm.² for serum globulin to 1.111×10^{-7} g./cm.² for serum albumin (Gorter 1934, 1935, Gorter and van Ormondt 1935; Schulman and Rideal 1933). With an average residue weight of 120, these densities give an area per residue ranging from 11.48 \AA^2 to 18.82 \AA^2 .

The cyclol hypothesis is now immediately applicable. A lamina formed by the polymerization of cyclol 6 gives a "film" with an area per residue of

(say) 10 \AA^2 . A less dense lamina can be built from polymers of cyclol 18 and cyclol 66 respectively, where the corresponding areas per residue are 13.34 \AA^2 and 16.36 \AA^2 respectively. Higher cyclols can polymerize to give even larger areas per residue.

We therefore suggest that the cyclol pattern (see fig. 9) is the basis of the structure of the unimolecular protein film, when it is approximately plane and one residue thick—in fact that these films are laminae comprising cyclol molecules or polymers of cyclol molecules. On this hypothesis we hazard the prediction that the upper limit of density of which such a film is capable without buckling, provided it is only one residue thick, is one residue per 10 \AA^2 , or taking the average molecular weight at 120, $1.98 \times 10^{-7} \text{ g./cm.}^2$. We would further assert that characteristic differences between the densities of films of different proteins indicate that the laminae in question are polymers of different cyclols.

Films of other proteins giving considerably higher densities have also been studied. Thus, for example, Hughes (1933) has obtained a gelatin film with density $2.5 \times 10^{-7} \text{ g./cm.}^2$. Gorter (1935, p. 427) has obtained for films of trypsin and pepsin densities of $(4.0 \text{ and } 7.7) \times 10^{-7} \text{ g./cm.}^2$, Rideal, Moss and Bate Smith (1935) a density of $(2 \text{ to } 3) \times 10^{-7} \text{ g./cm.}^2$ for a film of myosin. In these cases it is evidently impossible to interpret the films in terms of an approximately plane lamina one residue thick. They will, therefore, be further discussed in the sequel.

It may be remarked that the dorsal-ventral character of cyclol molecules of laminar type is in harmony with the surface chemistry of proteins, since the fact that proteins are capable of existing in the form of a film one residue thick implies that such a film must have one surface free from the hydrophobic groups in the side chains (and rings). The hypothesis further implies that the laminar molecules have a plentiful supply of the hydrophilic hydroxyls on both surfaces. It is of real interest that the uniform configuration of all the amino and imino acids obtained from proteins (which, as explained above, implies that these laminar molecules have one side free from side chains) is hereby brought into relation with the existence of protein films one residue thick, as is also the further implication that the laminar molecules carry a plentiful supply of hydroxyls on both sides.

The existence of these films is also significant in another aspect. A polypeptide chain lying on a water surface must necessarily arrange itself so that hydrophobic groups in the side chains (and rings) remain outside, clear of the water. A fully extended polypeptide chain cannot fulfil these conditions. It is geometrically necessary that it should curve round as in fig. 6. What is then more natural than to assume that, in the course of

freeing the hydrophobic groups of its side chains (and rings) from the water medium, the polypeptide falls into hexagonal folds, and in consequence C and N atoms approach so near to one another that new C—N links are formed? These protein films, one residue thick, may contain the secret of the genesis of the protein molecule.

6—LAMINAR MOLECULES AND MOLECULAR AGGREGATES

The dorsi-ventral character of the cyclol lamina suggests that lamina may be linked to lamina by superposition in two radically different ways.

The existence of the one and only "double" amino acid, cystine, here becomes of significance. Any cystine residue will from the start be a member of two laminae. The first way, therefore, of holding one lamina above another lamina is by means of cystine residues. It is possible that dicarboxylic and dibasic residues, which are theoretically capable of forming covalent links by condensation of the —COOH , H_2N -groups belonging to their side chains, may also play a part in holding lamina to lamina. Two cyclol molecules linked "front to front" in this way by means of the side chains will form one covalent unit.

The linking of lamina to lamina "back to back" brings in its train a host of interesting questions. It is the characteristic of these back surfaces that they carry a plentiful supply of hydroxyls, as many as one for every two residues in the polymer of cyclol 6. It is reasonable then to postulate linking between "back" surfaces by means of the hydroxyl groups.

It so happens that an investigation has recently been made (Bernal 1933, Bernal and Megaw 1935) as to the possible existence of hydroxyl links in many different classes of compounds. This "hydroxyl bond" was originally called the double hydrogen bond (Bernal and Fowler 1935), and the clue to its structure was provided by the study of ice (Megaw 1934). Further light was thrown upon it by a detailed analysis of the structure of aluminium hydroxide, in which hydroxyls play a part in holding together the composite layers of this crystal, there may also be a similar situation in the layer lattices of some of the clay minerals such as halloysite (Mehmel 1935). There is therefore a precedent for the suggestion that lamina may be held above lamina by means of attractions between hydroxyls. In the closely analogous case of the clay mineral, montmorillonite (Pauling 1930; Nargelschmidt 1934, Hofmann 1934*b*), such a structure is associated with a considerable capacity for hydration, an outstanding characteristic of many proteins. The structure of graphitic oxide (Hofmann 1934*a*) may also profitably be studied in this connexion.

The binding of lamina to lamina by means of hydroxyls is being considered in detail in another communication, with special reference to the problem of hydroxyl links in the polysaccharides (Haworth 1934, 1935). Attention will also be directed to the part played by the hydroxyl linkage in the energy balance mentioned in § 2 and to the implication that the instability of the laminar molecule (if indeed it is unstable) can be remedied by hydration or by the building of such laminae into laminate aggregates. For the moment it is sufficient to remark that a molecular aggregate in which one or more pairs of laminae are held together in this way will necessarily be sensitive to changes in the acidity of the medium. In particular, a sufficiently high pH will cause such an aggregate to dissociate into units, which are either single laminae or pairs of laminae joined together by cystine bridges or side chains in covalent links

The question of the hydroxy amino acids, serine, tyrosine, hydroxy glutamic acid (or hydroxy glutamine), hydroxy lysine, and the recently discovered α -amino β -hydroxy n -butyric acid (Womack and Rose 1935, Mc'oy, Meyer and Rose 1935) and the hydroxy imino acid, hydroxy proline, may also be mentioned here, since the hydroxyl groups on the side chains (or rings) may form hydroxyl bonds either with corresponding groups on other side chains or even with hydroxyls belonging to triazine hexagons

7—THE CLASSIFICATION OF PROTEINS

The cyclol hypothesis, restricted on this occasion to the case when the single cyclol molecule is approximately plane, and each lamina consists of a single cyclol molecule, suggests a classification of proteins on the following lines

First comes the single cyclol lamina. This can exist in a variety of forms each with its characteristic density which has already been discussed. These are the unimolecular protein films, one amino-acid residue thick.

Next comes the bilaminar molecule, in which the two single laminae are held together by cystine bridges or by side chains in covalent linkage.

With the present restrictions, these two types are the only true protein molecules, since they alone represent a covalent unit

When two or more of these covalent units are linked by means of their hydroxyls, we obtain a protein aggregate. By means of hydroxyl linkages, a molecular aggregate can thus be built which consists of as many laminae as we please. Thus a bilaminar molecular aggregate consists of two laminae linked "back to back", and a trilaminar molecular aggregate will consist of one lamina and one bilaminar molecule. A lamina built from four laminae

may comprise either two bilaminar molecules or one of these molecules and two laminae. A molecule built from five laminae will consist of two bilaminar molecules and a single lamina and so on.

The importance of the study of protein films is then apparent. A protein film, which is but one amino-acid residue thick, yet possesses the characteristic cyclol structure. Given any protein film we can discover by means of the density, or better the area per residue, whether or not it consists of a single lamina lying approximately in one plane. Such unilaminar films have already been the subject of much study. Several examples have already been referred to in an earlier paragraph. It was there suggested that a range of densities up to 1.98×10^{-2} g/cm.² indicates a single approximately plane lamina, distinctively different densities implying polymers of different cyclols. As was pointed out above, each type of polymer has not only a characteristic density but also characteristic lacunae. Roughly speaking, the lower the density the larger the lacunae. The geometry of these polymers in fact opens the way to an interesting field of research in which an attempt is made to study the penetrability of films of different proteins with a view to deducing the size of their lacunae.

Such researches will also find a field of applicability among protein films which are more than one amino-acid residue thick. The films of trypsin, pepsin, and myosin referred to above may be of this nature. A comprehensive study of the penetrability of laminar films will enable us to draw certain deductions from the penetrability of these laminate (or multilaminar) films, and so to check the deductions made from their densities.

The researches which we have suggested give a new interest to a large body of work on the penetrability of protein gels. Ruhland (1912), for example, examined the penetration of a large number of dyes into protein gels and grouped them according to their rate of penetration. He found that a dye such as methylene blue penetrated the gel as rapidly as a column of water, others penetrated more slowly and finally the colloidal dyes such as Bismarck brown and night blue were unable to penetrate the gel at all. The inference drawn was that the gel contained "capillaries" of a certain size.

These results may easily be interpreted on the present hypothesis. "The most important property of a gel", it has been said (Jordan Lloyd 1926 p. 223; 1920), "is its mechanical rigidity, combined with a remarkable lack of interference with the movements of the ordinary dissolved molecules of crystalloids." The cyclol hypothesis then suggests that in a gel there is an indefinitely extended structure, in which laminae are linked to laminae so that any three-dimensional space can be filled. Such a structure will possess

capillaries. The laminate aggregates formed by the superposition of laminae will possess capillaries of characteristic size depending upon the particular cyclols of which the individual laminae are composed. (Thus see fig. 12.) There will also be capillaries formed by the juxtaposition of laminate aggregates.

8—SVEDBERG'S RESULTS

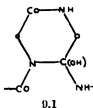
Reference has already been made to the fact that the present hypothesis implies that proteins fall into a few molecular weight classes, a conclusion which may throw light on certain of Svedberg's results. We now direct attention to the light thrown by the classification of proteins here adopted on another group of results established by Svedberg. Svedberg has shown that a number of different native proteins break up into smaller molecules with submultiple molecular weights (Svedberg 1930*b*). Thus amandin, stable from pH 4.3 to 10, at pH 12.2 splits up partly into molecules with one-sixth of the normal molecular weight. Edestin, stable from pH 5.5 to 10, at pH 11.3 breaks up partly into molecules with one-half and one-third of the normal molecular weight. Excelsin, stable from pH 5.5 to 10, is completely dissociated into molecules with one-sixth of the normal molecular weight at pH 11.9. Legumin, stable from pH 5 to 9, at pH 9.3 is partly dissociated into molecules of half the normal molecular weight. Three other proteins are also of interest in this connexion. *C* phycoeyan, *R* phycoeyan, and *R* phycoerythrin, with isoelectric points at pH 4.76, 4.5 and 4.25 respectively. At pH 6.8 the first is partly split up into molecules of half the molecular weight, at pH 12 it is completely dissociated into molecules with one-sixth of the normal molecular weight. The second is split up into molecules with half the molecular weight at pH 6.8. The third is partly split up into molecules with one-sixth the normal weight at pH 11.0.

On the present theory this dissociation of protein molecules into smaller molecules with submultiple molecular weights may be explained as due to the breakdown of linkage by means of hydroxyl bonds, consequent upon a too alkaline pH . The particular submultiples which occur may be regarded as affording evidence about the type of symmetry possessed by the laminae composing the molecular aggregates. Light could be thrown upon the situation by an investigation of the effect of a sufficiently alkaline pH on the dissociation of alcohols which at an appropriate pH exist in a highly associated form. The fact that a suspension of graphitic oxide, on the addition of alkali, divides into single layers is also suggestive (Hofmann 1935).

9—KERATIN AND THE FIBRE PROTEINS

This preliminary communication has necessarily been concerned with the more obvious deductions which can be made from the one and only assumption in the theory, namely the existence of double and triple peptide links in the protein molecule. We have reserved for detailed discussion in another publication the implications of the theory with respect to individual proteins, such as pepsin, insulin, and Bence Jones protein, for which the data are sufficiently precise to allow certain details of their structure to be predicted.

It is, however, but appropriate to mention the case of keratin, and in particular the "quasi-diketopiperazine" structure put forward for α -keratin by Astbury and Woods (1933), because the introduction of the linkage transformation (2.1) or of the prototropic transformation* (2.2) clears up a stereochemical anomaly in this structure. The structure of α -keratin as it now stands contains (using our nomenclature) a double peptide link.



With the introduction of keratin into our argument, the whole question of the structure of the fibre proteins presents itself for consideration. A simple extension of the cyclol hypothesis suggests a general structure for proteins of this type, which depends essentially upon a new interpretation of elasticity.

On this view proteins are elastic when the polypeptides in the molecules can take up a series of configurations which differ in the degree of cyclization. It at once follows that such a compound structure, in which many partially cyclized polypeptides are linked, need not (possibly even cannot) be dorsi-ventral, and the deductions as to the nature of laminate molecules therefore does not hold. The dorsi-ventrality which has so often been referred to in this communication is a property of one single cyclol molecule.

The further development of these ideas will appear in another publication.

* As suggested by F. C. Frank, see Astbury (1936).

10—THE CYCLOL PATTERN AND THE DENSITIES OF PROTEINS

It is essential that any theory of protein structure should satisfy the facts as to the densities of proteins.

We have already worked out some figures provided by the theory for the densities of protein films. The theory predicts that in no film can the area per residue be less than $B = 10 \text{ \AA}^2$. With an average residue weight of 120 this corresponds to a density of $1.98 \times 10^{-7} \text{ g/cm}^3$. The present theory predicts that this is the highest density possible, not that this density is ever in fact attained. It provides for consideration a set of different laminae with characteristically different areas per residue. The situation can best be explained by reference to the pattern made by the polycondensation of residues shown in fig. 14. The examples cited were: (1) the polymer of cyclol 6 shown in fig. 11 with the (minimum) area B per residue, (2) the polymer of cyclol 18 shown in fig. 12, and (3) the polymer of cyclol 66, with areas $4B/3$ and $18B/11$ per residue respectively. If now we refer to fig. 14, we see that (1) the polymer of cyclol 6 represents the polycondensation of the residues in which no residues are missing. On the other hand, in (2) 1 in 4 are missing and in (3) 7 out of 18. The pattern shown in fig. 14 represents the densest possible packing of the residues, and all the laminae under consideration are simply this pattern with certain proportions of the residues missing.

Evidently a lamina containing residues with long side chains may not represent the pattern of fig. 14 in its entirety. The densest of the laminae may perhaps only be formed by residues among which glycyl residues predominate. Similarly, a lamina containing a number of imino-acid residues must contain a corresponding number of single or double, as opposed to triple, peptide links: this means that certain numbers of residues in the pattern are missing and gives a correspondingly larger area per residue.

The theory therefore proposes for consideration a class of laminae in which the area per residue ranges from $B = 10 \text{ \AA}^2$ upwards, and it interprets a higher density as meaning that a film is more than one residue thick. Examples in which the area per residue is 13.34 \AA^2 and 16.36 \AA^2 respectively have been cited. The figures for the area per residue in films of gliadin, glutenin, egg albumin, zein, serum albumin and serum globulin range from 11.48 \AA^2 to 18.82 \AA^2 , if the average residue weight is 120.

These indications are for the moment sufficient to show that an analysis of the number of each type of residue in a protein film would enable us to go some way towards deciding between the various kinds of laminae (e.g.

between the examples (1), (2), and (3) cited above). It will be necessary to go into the stereochemistry of the various laminae in great detail. At present it seems difficult to see how a sufficient number of residues with long side chains and rings can be fitted into the denser laminae. The conclusions drawn earlier, that the film of serum albumin corresponds to a lamina with an area per residue intermediate between that characteristic of the polymers of cyclol 18 and of cyclol 66 respectively, point in the same direction.

The densities of laminate aggregates can also be worked out. We take the side chain spacing as 10 Å, as is customary (Astbury and Woods 1933), for the moment we assume that the hydroxyl link spacing from lamina to lamina is about 5.4 Å. The average distance between laminae, in a laminate aggregate in which these spacings alternate, will be about 7.7 Å. A laminate aggregate in which the single lamina is of type (1) would then give the volume per residue $7.7B = 77 \text{ Å}^3$. For (2) and (3) respectively these figures are 103 Å^3 and 126 Å^3 . If we take the average residue weight as 120, the densities will be 2.57, 1.93, 1.57 respectively in the three cases. There is no reason, on our hypothesis, to think that the maximum density is ever attained. Furthermore, cases in which the individual laminae are polymers of higher cyclols than cyclol 66 will give still lower densities.

The suggestion has been made that the fibre proteins may be made up of laminae which are no longer dorsal-ventral. In this case the distance between laminae will be 10 Å and the corresponding densities in the three cases will be 1.98, 1.48, 1.21.

These figures will be compared with the densities of the various proteins in order that more specific suggestions can be made as to the nature of the constituent laminae in the various cases. The average density for proteins in general is in the neighbourhood of 1.3. The hypothesis therefore allows the construction of laminar and laminate molecules and molecular aggregates with densities of the right order of magnitude.

11—CONCLUSIONS

So far as our initial researches are concerned there appears to be nothing inconsistent with the cyclol hypothesis in protein literature. The fact that it covers the four groups of data mentioned at the outset gives it the right to further consideration. We would, however, again draw special attention to two possible difficulties: the energy balance sheet, and the stereochemistry, particularly when residues with long side chains are involved.

It would increase our measure of confidence in the hypothesis if it were

possible to adduce definite experimental evidence. [Note added in proof 31 March 1937. A preliminary account is given in *Science* (1937) 85, 76, of a series of experiments by Langmuir, Schaefer and Wrinch which confirm the cyclol hypothesis at various points.] Happily the hypothesis is capable of statement in precise terms, and so suggests a number of crucial experiments belonging to many different fields: these will readily show either that it is false, or that it should be subjected to further experimental test.

Pending the test of direct experiment, we claim some measure of support for the hypothesis from the following considerations:

1. The theory gives an unforced significance to the two uniformities in the amino and imino acids obtained from proteins.

2. The theory reconciles the "polypeptide" and the "diketopiperazine" schools of thought in protein chemistry.

3. It suggests a class of structures which satisfy the experimental data with regard to protein films which are one residue thick, since these structures possess characteristically different densities ranging from one residue per 10\AA^2 downwards.

4. The theory derives from the single postulate of multiple peptide links a characteristic pattern which provides a common basis for the structure of all proteins. The existence of this common pattern explains why chemically different proteins yet share many properties in common; why, in fact, such a subject as protein chemistry exists at all. The immense number of ways in which the residues can be arranged in the common pattern show how this one class of substance can yet yield the variety required by physiology, biochemistry, and other domains of science concerned with the living organism, in particular by immuno- and enzyme-chemistry.

5. There is an easy accord between the theory and some of the results obtained by Svedberg.

I offer my thanks in connexion with the preparation of this paper to Professor Hirst, Professor Neville, Professor R. Robinson, F.R.S., W. T. Astbury, J. D. Bernal, D. M. Crowfoot, M. Dixon, H. R. Ing, D. Jordan Lloyd, H. S. Sobotka, and to Dr Warren Weaver and the Rockefeller Foundation. I am particularly indebted to M. F. Howson who was good enough to draw all the diagrams.

SUMMARY

A theory of the structure of the molecule of simple native protein, sufficiently precise to lend itself to experimental test by the methods of

organic, physical, enzyme and immunochemistry, is offered as a working hypothesis.

Designed to cover certain well established facts in protein chemistry, it suggests that the molecule may contain closed polypeptides and that the polypeptides are cyclized; i.e. they fall into a certain hexagonal ring structure. This hexagonal ring structure, the "cyclol" pattern, obtained by the cyclization of polypeptides, may be regarded more fundamentally as a surface polycondensation of amino and imino acid molecules. In it the polypeptide, the essential unit in the protein molecule on the classical theory, is replaced by the cyclized polypeptide; and correspondingly the peptide link, which on the classical theory joins one pair of C, N atoms, is replaced by multiple peptide links, in which two or three pairs of C, N atoms are involved. A series of cyclol laminar molecules are discussed with special reference to the unimolecular protein film. Attention is also directed to three-dimensional laminate (multilaminar) molecules and molecular aggregates, consisting of a number of single laminae superposed and held in position by side chains in covalent linkage and by hydroxyl links.

REFERENCES

- Abderhalden 1923 *Hoppe-Seyl. Z.* **128**, 119.
 Abderhalden and Kohn 1924a *Hoppe-Seyl. Z.* **134**, 121
 — — 1924b *Hoppe-Seyl. Z.* **139**, 147.
 — — 1924c *Hoppe-Seyl. Z.* **143**, 128
 Astbury 1936 *J. Text. Inst., Manchr* (in the Press).
 Astbury and Lomax 1935 *J. Chem. Soc.* p. 846.
 Astbury and Woods 1933 *Phil. Trans. A*, **232**, 333
 Bergmann 1935 *J. Biol. Chem.* **110**, 171
 Bergmann and others 1924 *Hoppe-Seyl. Z.* **140**, 128
 — 1925 *Hoppe-Seyl. Z.* **146**, 247.
 — 1926a *Liebigs Ann.* **448**, 38.
 — 1926b *Biochem. Z.* **177**, 1.
 Bernal 1931 *Z. Kristallogr.* **78**, 363
 — 1933 *Ann. Rep. Chem. Soc.* p. 401
 Bernal and Crowfoot 1934 *Nature, Lond.*, **135**, 794
 Bernal and Fowler 1935 *J. Chem. Phys.* **1**, 515.
 Bernal and Megaw 1935 *Proc. Roy. Soc. A*, **151**, 384
 Clough 1918 *J. Chem. Soc.* **113**, 526.
 Cohn 1928 *Physiol. Rev.* **8**, 349.
 Crowfoot 1935 *Nature, Lond.*, **135**, 591.
 Fischer 1906 *Ber. dtsch. chem. Ges.* **39**, 607.
 Fischer and Raske 1907 *Ber. dtsch. chem. Ges.* **40**, 3717
 — — 1908 *Ber. dtsch. chem. Ges.* **41**, 893.
 Gorter 1934 *Amer. J. Dis. Child.* **47**, 945
 — 1935 *J. Gen. Physiol.* **18**, 421.
 Gorter and van Ormondt 1935 *Biochem. J.* **29**, 48

- Haworth 1934 *J. Soc. Dy. Col., Bradford*, p. 16.
— 1935 *Rep. Brit. Ass.* p. 31.
Hofmann 1934a *Liebigs Ann.* 510, 1.
— 1934b *Kolloidschr.* 69, 120.
— 1935 *Kolloidschr.* 69, 351.
Hughes 1933 *Trans. Faraday Soc.* 29, 211.
Jordan Lloyd 1920 *Biochem. J.* 14, 147, 584.
— 1926. "The Chemistry of the Proteins."
— 1932 *Biol. Rev.* 7, 254.
Karrer and Kaase 1919 *Helv. chim. Acta*, 2, 436.
— — 1920 *Helv. chim. Acta*, 3, 244.
McCoy, Meyer and Rose 1935 *J. Biol. Chem.* 112, 283.
Megaw 1934 *Z. Krystallogr.* 87, 185.
Mehmel 1935 *Z. Krystallogr.* 90, 35.
Nargelschmidt 1934 *Z. Krystallogr.* 87, 120.
Pauling 1930 *Proc. Nat. Acad. Sci., Wash.*, 16, 123, 578.
Rideal, Moss and Bate Smith 1935 *Nature, Lond.*, 136, 280.
Ruhland 1912 *Jb. wiss. Bot.* 51, 376.
Schmper 1881 *Z. Krystallogr.* 5, 131.
Schulman and Rideal 1933 *Biochem. J.* 27, 1581.
Svedberg 1930a *Kolloidschr.* 51, 10.
— 1930b *Trans. Faraday Soc.* 26, 737.
— 1933 *Chem. Rev.* 14, 1.
— 1934 *Science*, 79, 327.
Waser and Brauchli 1924 *Helv. chim. Acta*, 7, 740.
Wornack and Rose 1935 *J. Biol. Chem.* 112, 275.
Wrinch 1936a *Nature, Lond.*, 137, 411.
— 1936b *Nature, Lond.*, 138, 241.
Wrinch and Jordan Lloyd 1936 *Nature, Lond.*, 138, 758.
Wyckoff and Corey 1934 *Z. Krystallogr.* 89, 462.
-

The Relationship between Viscosity, Elasticity and Plastic Strength of a Soft Material as Illustrated by some Mechanical Properties of Flour Dough

IV—The Separate Contributions of Gluten and Starch

BY R. K. SCHOFIELD AND G. W. SCOTT BLAIR

(Communicated by Sir John Russell, F.R.S.—Received 21 September 1936—
Revised 15 January 1937)

In the earlier papers of this series (Schofield and Scott Blair 1932, 1933*a*, 1933*b*) the endeavour has been to give a quantitative description of the behaviour of flour dough under stress. Use was made of the equation

$$\frac{de}{dt} = \left(\frac{1}{n} \cdot \frac{dS}{dt} - \frac{dx}{dt} \right) + \frac{1}{\eta} \cdot S,$$

which is the expression originally put forward by Maxwell with the addition of $-dx/dt$ to take account of elastic after-effect. In this equation de/dt represents the rate of elongation of a cylinder of dough, and S the shearing stress which is one-third the longitudinal stress per unit area. The equation serves to define n , the modulus of rigidity, and η the viscosity, and enables these to be evaluated from experimental observations of e and S .

The behaviour of flour dough was shown to be consistent with the equation

$$t_r = \eta/n$$

for the relaxation time, and evidence has since been presented by Halton and Scott Blair (1936) to show that t_r is probably the most important single quantity determining the baking quality of a flour dough.

Valuable as this method of formulation has undoubtedly been, it does not throw much light on the physical mechanisms at work, its virtue being that it can be applied to any material that can be stretched. By its use a remarkable *formal* resemblance was demonstrated between the behaviour of flour dough and soft metals: both show work hardening, elastic hysteresis and elastic after-effect. It cannot be argued from this, however, that the physical mechanism is the same. To study this aspect of the problem other avenues of approach must be sought.

THE MECHANISM OF WORK HARDENING IN FLOUR DOUGH

The Gluten Network—An investigation of the behaviour of dough cylinders when extended to many times their original length has provided a clue to the explanation of work hardening

The apparatus was a simplified form of that already described. The cylinder of dough originally 5 cm. long and about 0.7 cm. in diameter was floated straight out of the "moulding gun" on to the mercury bath which was provided with a lid lined with wet felt to prevent the dough from drying. Threads of sewing cotton were attached, one to each end of the cylinder, by means of small pieces of cork to which dough adheres readily. One thread was attached to a stress indicator, reading up to about 7000 dynes,* of stouter design than that used previously, and the other to a winch geared to a synchronous motor which enabled it to be moved steadily at a rate of 0.045 cm./sec.

In the first series of experiments represented by figs. 1, 5 and 8, the stress was recorded at intervals during the slow extension at the end of which the cotton was released from the winch. The dough cylinder was then allowed 5 min. for free contraction before the slow extension was repeated. In these figures the shearing stress per unit area (dynes/cm.²) is plotted against $\log_e l/l_0$, l being the length at the moment in question and l_0 the initial length (5 cm.).

The fact that the length of a dough cylinder which has completed its elastic recovery after being stretched always exceeds the original length, is most simply explained by supposing that the elastic elements in the dough are insecurely attached to each other. It might be expected that a general slippage would occur when a critical stress had been reached. This, however, is not the case in a normal dough. Fig. 1 shows clearly that although considerable flow took place in the first extension, very much less occurred during the subsequent extensions although somewhat higher stresses were recorded. The upward curvature at the higher strains which is such a pronounced feature of the curves is just discernible in fig. 7 of the paper by Schofield and Scott Blair (1933*b*) which was only carried to a strain $\log_e l/l_0 = 0.7$.

The general nature of the process appears to be something like that indicated in fig. 2, which represents the behaviour of six springs, three of which are securely linked at P , the other three at Q , while they are insecurely linked in pairs at R , S and T . The springs linked at S are only about half the length of those linked at R and T . Consequently, if each insecure link

* 15 cm. at 0.455 g./cm. = 6700 dynes.

will stand the same maximum stress, the link at *S* will snap before those at *R* and *T*. If only a small stress is applied (fig. 2*b*) all the links will hold, and the system when released will recover to the unstrained length (fig. 2*a*). But, once *S* has snapped (fig. 2*c*) the system will not recover to its length in

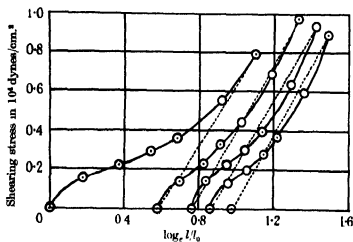


FIG. 1

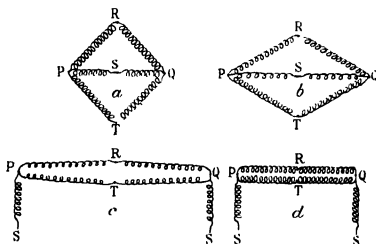


FIG. 2

fig. 2*a* but only to the condition shown in fig. 2*d*. The system has undergone a permanent elongation and has therefore "flowed". It must be remembered that the dough is to be pictured as formed of a very great many such systems. The more junctions that have broken, the higher must the applied stress be before there is any more breaking of junctions. Work hardening

may be envisaged either as a rise in yield value, or in viscosity, but in either case it is clear that the progressive snapping of individual links has produced hardening.

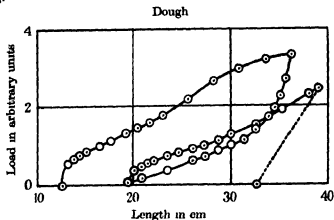


FIG. 3

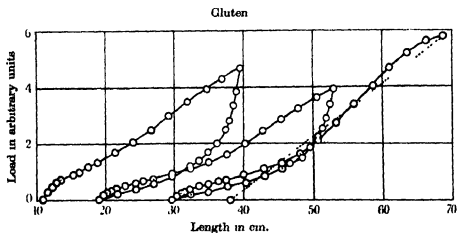


FIG. 4

For more easy comparison with the model of fig. 2 the behaviour of a dough cylinder and of a strip of washed gluten* are shown in figs. 3 and 4 by plotting simply the load against the length. In these experiments, when

* Gluten test-pieces are prepared in the following way: A small ball of dough is kneaded continuously under the tap until all the starch has been washed away, and the wash water is clear. The mass of wet gluten, which consists largely of protein and water, is given coherence by further kneading, and a small strip of approximately rectangular cross-section is cut with a razor. When extended on the trough of the extensometer, this loses its angles, and recovers to the shape of a slightly irregular cylinder whose diameter can be measured with a fair degree of accuracy.

the cylinder had been extended as far as was considered safe, the synchronous motor was reversed and the cylinder was allowed to contract at the same rate as it had been extended.

It will be seen that for the greater part of their length the loading branches of the curves bend upwards. This is the behaviour to be expected of a system of springs irregularly assembled, some of which do not come under stress at all until a certain amount of elongation has taken place. Such a curvature would also appear where the springs are approaching their greatest possible extension.

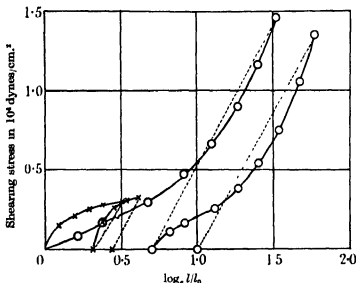


FIG. 5—○ Control. × HCl treated.

It is evident from the close resemblance between figs. 3 and 4 that the elastic structure in the dough is the gluten. It is, therefore, not unreasonable, in view of the work of Astbury (1933) on protein fibres, to suggest that the branched protein chains of the gluten are the springs securely fastened together, and that the insecure links are made by the electrostatic attraction between oppositely charged groups of neighbouring molecules.

Support for this view was obtained from an experiment in which hydrochloric acid was added to a dough slightly in excess of the amount needed to convert all the COO^- groups into COOH . The maximum shearing stress that this dough would stand without tearing was about 0.35×10^4 dynes/cm.², whereas the control dough made up to the same moisture content but without the addition of acid easily withstood 1.5×10^4 dynes/cm.². The curves are shown in fig. 5.

The effect of 10 min. drastic remixing of a dough is shown in fig. 6, from which it will be seen that this treatment caused the cylinder to flow out under a relatively low stress. The normal condition was, however, largely restored by allowing the remixed dough to stand for 2 hr.

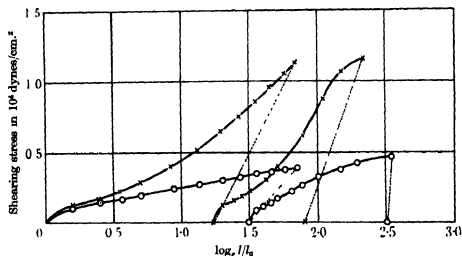


FIG. 6—O Immediately after drastic remixing. x After resting further 2 hr.

Hence it was evident that a considerable time is required for the elastic structure in a dough to become linked up. In order to investigate this point further, cylinders were prepared from a freshly made dough and kept on the mercury bath under small waterproof covers for varying lengths of time before being tested. The cylinders were extended for 5 min. by which time they had been stretched by $\log_e l/l_0 = 1.28$, and the stress was recorded. The stress was then released and $\log_e l/l_0$ again determined after 5 min free contraction. They were again extended, this time until rupture occurred, and their greatest length was measured. Judged by the length at rupture

TABLE I

	Age (hours)	...	0.33	1.0	2.0	2.75	4.5	7.5
First extension								
Shearing stress for $\log_e l/l_0 = 1.28$ in			1.22	0.71	0.61	0.65	0.62	0.58
10^4 dynes/cm. ²								
$\log_e l/l_0$ after recovery			0.69	0.78	0.77	0.77	0.87	0.91
Second extension								
Length at rupture (cm.)			20	33	29	34	32	32

there is little further build-up of structure after 1 hr., but it will be seen that the flow during the first extension, as measured by $\log_e l/l_0$ after recovery,

continues to increase up to 7 hr. notwithstanding the fall in shearing stress. This may be due to water released by syneresis of the gluten diluting the starch paste, and thereby reducing its viscosity (*vide infra*).

The Influence of the Starch—In comparing the curves for dough and gluten shown in figs. 3 and 4, one is struck by their very close similarity. In fact, apart from a tendency for the loops to be wider in the case of the dough, there is only one noticeable difference. At the lowest strains, the stress rises very steeply in the case of dough to what almost is a yield value, whereas with gluten there is considerable deformation even at the lowest stresses. The only difference between gluten and dough is that the latter contains starch, whereas in the former this ingredient has been removed. Starch paste has an anomalous viscosity, i.e. one that is higher the lower the shearing stress, and so would be expected to influence the behaviour of the dough more at low than at high stresses.

Some additional experiments in which equilibrium was approached first by a simple recovery from a small load, and secondly after having momentarily overloaded the sample, showed that the loops in figs. 3 and 4 are mainly due to true hysteresis and not to any great extent to elastic after-effect: they would have appeared even if the cycles had been carried out very much more slowly.

It was also found that under very small loads applied in the range where no permanent deformation occurs, dough shows an elastic fatigue on repeated loading, but that it slowly recovers on resting. This is evidently a type of thixotropic behaviour.

SUMMARY

Experiments are described which support the view that in a flour dough the gluten forms an elastic network which dominates the mechanical behaviour. It appears that when a cylinder of dough is first stretched some of the links in the network are ruptured since it will not return to its original length. Enough remain unbroken, however, for a continuity of structure to be preserved until the cylinder has been extended to five or six times its original length. The "work hardening" of dough is thus accounted for. The elastic network does not establish itself at once, but continues to build up for some time after the dough is mixed. Its strength is greatly reduced by drastic remixing of the dough but is largely recovered on further standing. The addition of hydrochloric acid in slight excess of the acid binding capacity destroys the strength of the network. This shows that the electro-

static attraction between oppositely charged groups in neighbouring molecules is an important factor in the strength of the gluten network.

The upward bend of the reloading curve up to the point where flow (i.e. the rupture of further links) occurs is probably mainly due to the irregularity of assembly of the elastic members, but may also indicate that individual chains are approaching the limit to which they can be extended.

Evidence has been obtained that the starch paste penetrating the gluten network has a "yield value", in consequence of which there is elastic hysteresis even when the cycle is carried out slowly enough to avoid elastic after-effect.

REFERENCES

- Asbury, W. T. 1933 "Fundamentals of Fibre Structure." Oxford Clarendon Press
Halton, P. and Scott Blair, G. W. 1936 *J. Phys. Chem.* **40**, 561
Schofield, R. K. and Scott Blair, G. W. 1932 *Proc. Roy. Soc. A*, **138**, 707.
— — 1933^a *Proc. Roy. Soc. A*, **139**, 557
— — 1933^b *Proc. Roy. Soc. A*, **141**, 72

The Quantum Theory of Atomic Polarization I—Polarization by a Uniform Field

By R. A. BUCKINGHAM, B.A., *Queen's University, Belfast*

(Communicated by J. E. Lennard-Jones, F.R.S.—Received 29 December 1936)

1—INTRODUCTION

Two problems of atomic energy, the energy of polarization of an atom in a plane electrostatic field and the energy of interaction, or van der Waals energy, of two distant atoms, are particularly suited to attack by approximate methods. In each, the disturbing field, whether that of the static field or of the distant atom, is small in comparison with the internal fields acting on the electrons of the atomic system, and so the standard methods by which quantum mechanics deal with small disturbances, namely the perturbation and variation methods, can be and have been successfully applied to these problems. Though the perturbation method, strictly applied, is the more accurate, since it takes into account the possible excited states of the

system, its usefulness is restricted to atoms of simple structure, and having relatively simple wave functions. Generally speaking the variation methods, which require a knowledge of the unperturbed state of the system only, are more suitable for larger atoms; ignorance of the excited states is largely compensated for by expressing the perturbed wave functions in terms of parameters, which are then chosen to make the total energy a minimum.

As one would expect, the atoms of hydrogen and helium have been studied most fully, the perturbation theory being used by Wang, Eisenschitz and London, and Lennard-Jones, and variation methods by Atanasoff, Hassé, Slater and Kirkwood, Pauling and Beach.† Owing to the first order Stark effect, the calculated polarizability of the hydrogen atom cannot be compared with experiment, but the calculated value for helium agrees well in most cases with that observed. The van der Waals energy of two hydrogen atoms given by Pauling and Beach includes accurate values not only of the usual dipole-dipole term, which varies as $1/R^6$, but also of the dipole-quadrupole and quadrupole-quadrupole terms, varying as $1/R^8$, $1/R^{10}$ respectively.

The most accurate calculation of the dipole-dipole term for two helium atoms is probably that of Hassé (1930, 1931), and the higher order terms have been estimated by Margenau (1931).

The treatment of larger atoms has been largely devoted to obtaining a relation between the dipole-dipole constant c (the van der Waals energy being $-c/R^6$) and the polarizability α . It was started by London (1930), who deduced that for two similar atoms

$$c = (\text{const.}) I \alpha^2,$$

where I is the ionization potential. The use of the variation method by Slater and Kirkwood (1931) suggested that

$$c = (\text{const.}) N \alpha^4,$$

where N is the number of electrons in the outer shell, and though their work was restricted mainly to hydrogenic wave functions and functions without nodes, the method was later modified by Kirkwood (1932) to include any type of wave function. A different line of approach was followed by Vinti (1932), who used the Kuhn-Reiche sum rule to simplify the second order terms of the perturbation theory (the first order terms vanish in these problems), showing at the same time that the assumptions made were

† For the treatment of H: Wang (1927); Eisenschitz and London (1930); Lennard-Jones (1930a); Pauling and Beach (1935). For H and He, Hassé (1930, 1931), Slater and Kirkwood (1931). For He: Atanasoff (1930)

implied also in Kirkwood's variation method. Using the same idea, Hellmann (1935) was able to extend the results of Vinti and Kirkwood, by treating the electrons of each atom not as a single group but as a number of shells, which in the variation method corresponds to the use of several parameters instead of one, and by including the Pauli Principle.

In this paper, Kirkwood's variation method is used throughout, with one or more parameters and with inclusion of electron exchange, and formulae derived for the polarizability of atoms with any number of electrons. The results are somewhat more general than those of Hellmann. Calculations are then made with the available wave functions of the rare gas atoms and alkali ions, and the results compared with experiment as far as possible. In a second paper, the van der Waals energy of atoms will be considered.

2—ENERGY OF PERTURBATION

We shall begin by deriving an expression for the increase in energy of an atom containing N electrons when subject to a small perturbing field. We shall assume that the electronic system is in a non-degenerate state which can be adequately represented by a wave function involving the space and spin co-ordinates of the electrons only. Thus if the unperturbed wave function is Ψ , and its complex conjugate is Ψ^* , the energy of the system is given by

$$E = \int \Psi^* H \Psi d\tau / \int \Psi^* \Psi d\tau, \quad (2.1)$$

where

$$H = \sum_{p=1}^N \left\{ -\frac{1}{2} \nabla_p^2 + V_p \right\},$$

V_p being the potential energy of the p th electron, expressed as a function of the space co-ordinates of the p th electron only, and ∇_p^2 the Laplacian operator for the same electron. The integration is carried out over the co-ordinate space of every electron.

When an external field is applied to the system, let us suppose V_p to be increased by a small amount v_p to $V_p + v_p$. H is therefore increased to $H + v$, where $v = \sum v_p$, and at the same time the function Ψ is changed to $\Psi + \Phi$, where $\Phi^* \Phi$ is of the order of magnitude of v^2 . The energy of the perturbed state is then

$$E + \bar{h} = \frac{\int (\Psi^* + \Phi^*) (H + v) (\Psi + \Phi) d\tau}{\int (\Psi^* + \Phi^*) (\Psi + \Phi) d\tau}, \quad (2.2)$$

\bar{h} being the energy of perturbation we wish to evaluate. Now in the problems to be discussed later, v and $\Psi^*\Phi$ are odd functions† of the electron co-ordinates, and a number of integrals in (2.2) vanish, including

$$\int \Psi^* v \Psi d\tau, \quad \int \Phi^* v \Phi d\tau, \quad \int \Psi^* \Phi d\tau, \quad \int \Phi^* \Psi d\tau.$$

Hence

$$E + \bar{h} = \frac{1}{a+b} \left[\int (\Psi^* \Phi + \Phi^* \Psi) v d\tau + \int (\Psi^* + \Phi^*) H (\Psi + \Phi) d\tau \right], \quad (2.3)$$

where

$$a = \int \Psi^* \Psi d\tau, \quad b = \int \Phi^* \Phi d\tau.$$

By transforming the integral containing ∇_p^2 by Green's Theorem, so that the surface integral vanishes over the infinite boundary, and by neglecting terms in v^4 and higher powers of v , and also terms of the form

$$a \int \Phi^* V_p \Phi d\tau - b \int \Psi^* V_p \Psi d\tau, \quad (2.4)$$

we find that

$$\begin{aligned} \bar{h} = & \frac{1}{a} \int (\Psi^* \Phi + \Phi^* \Psi) v d\tau \\ & + \frac{1}{2a} \sum_p \left[\int \text{grad}_p \Phi^* \cdot \text{grad}_p \Phi d\tau - \frac{b}{a} \int \text{grad}_p \Psi^* \cdot \text{grad}_p \Psi d\tau \right]. \end{aligned} \quad (2.5)$$

The expression for \bar{h} given by Kirkwood and Hellmann is slightly different from (2.5), but follows immediately from it when Φ is replaced by Ψu , provided that we can neglect terms of the type

$$\int \text{grad}_p \Psi^* \cdot \text{grad}_p u^2 \Psi d\tau - \frac{1}{a} \int \Psi^* u^2 \Psi d\tau \int \text{grad}_p \Psi^* \cdot \text{grad}_p \Psi d\tau. \quad (2.6)$$

We then obtain

$$\bar{h} = \frac{1}{a} \int \Psi^* \{ 2uv + \frac{1}{2} \sum_p (\text{grad}_p u^2) \} \Psi d\tau. \quad (2.7)$$

Both (2.5) and (2.7) will be used in the applications which follow, and we shall assume that the terms of which (2.4) and (2.6) are typical are small‡ in comparison with other terms of the order of v^3 .

† The sign of the function is changed if the co-ordinates of any electron are reversed in sign.

‡ It can be shown that, when Ψ is real and satisfies the Schrödinger equation for the atom exactly, Kirkwood's result (2.7) holds without approximation. This suggests that when Φ cannot be written in the form Ψu , the neglect of the terms corresponding to (2.4) and (2.6) is in fact justified. It is intended to examine these terms in a later paper.

3—POLARIZATION BY A UNIFORM FIELD

The results of § 2 will first be applied to the increase in energy of an atom in a uniform plane electrostatic field. Let us assume that the wave function of the atom when the electrostatic field is absent, is the determinant

$$\Psi = \begin{vmatrix} \psi_{a1} \psi_{a2} & \dots & \psi_{aN} \\ \psi_{\beta 1} \psi_{\beta 2} & \dots & \psi_{\beta N} \\ \vdots & & \vdots \\ \psi_{r1} \psi_{r2} & \dots & \psi_{rN} \end{vmatrix}, \quad (3.1)$$

where $\psi_{a\mu}$, $\psi_{\beta\mu}$, ..., $\psi_{r\mu}$ are the wave functions of the occupied electronic states of the atom, the Greek suffix being an abbreviation for the four quantum numbers n, l, m, s , of the state, and the italic suffix indicating the electron whose co-ordinates are inserted. The ψ 's are assumed to include the spin function as a factor, to be orthogonal and each to be normalized to unity. If Ψ^* is the conjugate complex of Ψ then

$$a = \int \Psi^* \Psi d\tau = N!$$

if the integration is carried out over the co-ordinate space of every electron. Ψ also satisfies the Pauli Exclusion Principle and permits any pair of electrons to exchange places freely.

When the atom is subjected to an external field the ψ 's are changed. It is convenient to express the perturbed function corresponding to $\psi_{\rho p}$ which contains the co-ordinates of the p th electron, in the form

$$\psi_{\rho p} (1 + f(r_p) v_p)$$

where v_p is the increase in the potential energy of the p th electron and $f(r_p)$ is a polynomial function of its radial co-ordinate. As a first approximation we shall take $f(r_p) = \lambda$, a constant for all values of ρ and p . The perturbed wave function of the whole atom is then obtained by replacing $\psi_{\rho p}$ in the determinant Ψ by $\psi_{\rho p}(1 + \lambda v_p)$ and therefore, if only linear terms in v_p are retained, it is

$$\Psi(1 + \lambda v),$$

where $v = \sum v_p$. This function is likewise antisymmetrical in the electron co-ordinates, and so satisfies the Pauli Principle.

The energy of perturbation is now given by (2.7); inserting $a = N!$, $u = \lambda v$, we have

$$\bar{h} = \frac{1}{N!} \int \Psi^* \{ 2\lambda v^2 + \frac{1}{2} \lambda^2 \sum_p (\text{grad}_p v)^2 \} \Psi d\tau. \quad (3.2)$$

Let \mathbf{r}_p be the position vector of the p th electron, and \mathbf{F} a constant vector representing the electrostatic field. Then in atomic units, the field perturbing the p th electron is $v_p = -\mathbf{F} \cdot \mathbf{r}_p$ and so

$$v = - \sum_{p=1}^N (\mathbf{F} \cdot \mathbf{r}_p). \quad (3.3)$$

With this expression for v , (3.2) can easily be reduced to a sum of simple integrals involving one electron wave functions only. As this part of the process is long it is omitted here, though further details of this and similar reductions which arise later will be found in an appendix. We find that, after averaging for all orientations of the atom relative to the direction of the external field,

$$\bar{h} = F^2 (\frac{1}{3} \lambda \bar{R}^2 + \frac{1}{2} N \lambda^2), \quad (3.4)$$

where

$$\bar{R}^2 = \sum_p (R^2)_p$$

and $(R^2)_p = (r^2)_{pp} - \sum_{\sigma \neq \rho} \{ (x)_{\sigma\rho}(x)_{\rho\sigma} + (y)_{\sigma\rho}(y)_{\rho\sigma} + (z)_{\sigma\rho}(z)_{\rho\sigma} \}.$

The quantity $(r^2)_{pp}$ is the average value of r^2 with respect to the charge density $\psi_p^* \psi_p$, i.e.

$$(r^2)_{pp} = \int \psi_p^* r^2 \psi_p d\tau.$$

It is independent of the electron considered and so the electronic suffix is omitted. The other terms in $(R^2)_p$ are exchange integrals, of which

$$(x)_{\sigma\rho} = \int \psi_\sigma^* x \psi_\rho d\tau,$$

is typical. These integrals are discussed in detail in § 3a.

We see that expression (3.4) for \bar{h} contains the unknown parameter λ , which must be chosen to make the energy a minimum. Accordingly we introduce the condition $d\bar{h}/d\lambda = 0$ which gives $\lambda = -2\bar{R}^2/3N$ and

$$\bar{h}_{\min} = -\frac{2}{9} \frac{F^2}{N} (\bar{R}^2)^2$$

The quantity of practical importance is the atomic polarizability α defined by $\bar{h}_{\min.} = -\frac{1}{2} \alpha F^2$ and therefore given by

$$\alpha = \frac{4}{9N} (\bar{R}^2)^2 \quad (3.5)$$

in atomic units.

In deriving (3.5) it was assumed that each wave function ψ_{pp} is perturbed in the same degree, determined by the parameter λ . This is unlikely

to be accurate, if only because electrons in the inner shells are shielded by those in the outer from a perturbing field acting from without, and so we will consider the effect of using a different parameter for each electronic wave function. Thus let us now assume the perturbed form of $\psi_{\rho\rho}$ to be $\psi_{\rho\rho}(1 + \lambda_{\rho}v_p)$ where λ_{ρ} is a different constant for each state ρ from α to ν . The perturbed wave function of the atom, obtained by replacing $\psi_{\rho\rho}$ in Ψ (3.1) by $\Psi_{\rho\rho}(1 + \lambda_{\rho}v_p)$ now becomes, when only linear terms in v_p are retained,

$$\Psi + \Phi,$$

where $\Phi = \sum_{\rho} \lambda_{\rho} \sum_{\rho} \psi_{\rho\rho} D_{\rho\rho} v_p$, $D_{\rho\rho}$ being the first minor of $\psi_{\rho\rho}$ in the determinant Ψ . Φ is no longer a simple multiple of Ψ , and so equation (2.5) for the energy of perturbation must be used instead of (2.7). However, this does not affect the accuracy of the final result, as in the course of reduction to simple integrals we neglect certain terms comparable with (2.6). After averaging for all orientations of the atom as before, we obtain in place of (3.4)

$$\bar{h} = F^2 \sum_{\rho} \left\{ \frac{2}{3} \lambda_{\rho} (R^2)_{\rho} + \frac{1}{2} \lambda_{\rho}^2 \right\}, \quad (3.6)$$

and when the parameters are chosen to make \bar{h} a minimum,

$$\bar{h}_{\min.} = -\frac{2}{3} F^2 \sum_{\rho} (R^2)_{\rho}^2,$$

and

$$\alpha = \frac{2}{3} \sum_{\rho} (R^2)_{\rho}^2, \quad (3.7)$$

where the summation for ρ includes all the occupied electronic states.†

As a further refinement we shall take the perturbed form of $\psi_{\rho\rho}$ to be‡

$$\psi_{\rho\rho} \{1 + \lambda_{\rho} v_p (1 + \mu_{\rho} r_p)\},$$

† By applying the Kuhn-Reiche sum rule to the second order terms of the perturbation theory, Hellmann (1935) obtains the following expressions for α :

$$\alpha = 4 \sum_i (z^2)_{ii} \left\{ (z^2)_{ii} - \sum_{k \neq i} |(z)_{ik}|^2 \right\}.$$

Different orientations of the atom relative to the external field are not considered. When written in the same way (3.7) becomes

$$\alpha = 4 \sum_i \left\{ (z^2)_{ii} - \sum_{k \neq i} |(z)_{ik}|^2 \right\}^2.$$

It is not clear why the two results should differ.

‡ In dealing with He, Haseé (1930) showed that for the introduction of the linear term of the polynomial $f(r_p)$ improved the calculated value of the polarizability considerably. The higher powers of r_p were found to have a negligible effect.

where μ_ρ , like λ_ρ , is a different constant for each wave function. The analysis is scarcely affected and yields the result

$$\bar{h} = F^2 \sum_{\rho} \left[\frac{1}{3} \lambda_{\rho} \{ (R^2)_{\rho} + \mu_{\rho} (R^2)_{\rho} \} + \frac{1}{2} \lambda_{\rho}^2 \{ 1 + \frac{2}{3} \mu_{\rho} (r)_{\rho\rho} + 2 \mu_{\rho}^2 (r^2)_{\rho\rho} \} \right], \quad (3.8)$$

where $(R^2)_{\rho} = (r^2)_{\rho\rho} - \sum_{\sigma \neq \rho} \{ (rx)_{\sigma\rho} (x)_{\rho\sigma} + (ry)_{\sigma\rho} (y)_{\rho\sigma} + (rz)_{\sigma\rho} (z)_{\rho\sigma} \}$

and $(R^2)_{\rho}$ is the same as in (3.4). The minimum value of \bar{h} gives for the polarizability

$$\alpha = \frac{4}{9} \sum_{\rho} (R^2)_{\rho}^2 (1 + \Delta_{\rho}), \quad (3.9)$$

$$1 + \Delta_{\rho} = \frac{1 + \frac{(R^2)_{\rho}^2}{2(r^2)_{\rho\rho}(R^2)_{\rho}^2} - \frac{4}{3} \frac{(r)_{\rho\rho}(R^2)_{\rho}}{(r^2)_{\rho\rho}(R^2)_{\rho}}}{1 - \frac{8}{9} \frac{(r)_{\rho\rho}^2}{(r^2)_{\rho\rho}}}. \quad (3.10)$$

The quantity Δ_{ρ} , which may be regarded as a correction to the much simpler formula (3.7), is shown in § 4b to be important for the lighter atoms.

3a—Evaluation of Integrals

The functions $(R^2)_{\rho}$ and $(R^2)_{\rho}$ which occur in formulae (3.5), (3.7) and (3.10) for the atomic polarizability can easily be evaluated once the electronic wave functions are known. We have defined

$$(R^2)_{\rho} = (r^2)_{\rho\rho} - \sum_{\rho' \neq \rho} \{ (x)_{\rho'\rho} (x)_{\rho\rho'} + (y)_{\rho'\rho} (y)_{\rho\rho'} + (z)_{\rho'\rho} (z)_{\rho\rho'} \},$$

in which ρ, ρ' are written shortly for the quantum numbers n, l, m, s , and n', l', m', s' , respectively and

$$(r^2)_{\rho\rho} = \int \psi_{\rho}^* r^2 \psi_{\rho} d\tau, \quad (x)_{\rho\rho'} = \int \psi_{\rho}^* x \psi_{\rho'} d\tau, \text{ etc.}$$

ψ_{ρ} is usually expressed in spherical polar co-ordinates in the form $(2\pi)^{-1/2} R(nl|r) P_l^m(\cos \theta) e^{im\phi} \times \text{spin function}$, where $R(nl|r)$ is the radial wave function, normalized to unity, and $P_l^m(\cos \theta)$ the associated Legendre function. The angular and spin factors of $(r^2)_{\rho\rho}$, $(x)_{\rho\rho'}$, $(y)_{\rho\rho'}$ and $(z)_{\rho\rho'}$ can be integrated at once, and incidentally $(z)_{\rho\rho'}$ vanishes unless the quantum numbers of ψ_{ρ} and $\psi_{\rho'}$ satisfy

$$|l' - l| = 1, \quad m' = m, \quad s' = s;$$

likewise $(x)_{\rho\rho'}$, $(y)_{\rho\rho'}$ vanish unless

$$|l' - l| = 1, \quad |m' - m| = 1, \quad s' = s.$$

The integration for r introduces integrals of the types

$$(r^2)_{nl, nl} = \int_0^\infty r^2 [P(nl|r)]^2 dr, \quad (r)_{nl, n'l'} = \int_0^\infty r P(nl|r) P(n'l'|r) dr,$$

where $P(nl|r) = r \cdot R(nl|r)$ and we find eventually that for an atom containing complete (n, l) subgroups,

$$(R^2)_{nlms} = (r^2)_{nl, nl} - \frac{1}{2(2l+1)} \sigma_{nl}, \quad (3.11)$$

where σ_{nl} is a function whose form depends upon the azimuthal quantum number l , but is independent of the magnetic and spin quantum numbers. Thus, for an atom containing complete s, p, d and f shells:

$$\left. \begin{aligned} \sigma_{n0} &= 2 \sum_{n'} (r)_{n'l, n0}^2, \\ \sigma_{n1} &= 2 \sum_{n'} \{ (r)_{n'0, n1}^2 + 2(r)_{n'2, n1}^2 \}, \\ \sigma_{n2} &= 2 \sum_{n'} \{ 2(r)_{n'1, n2}^2 + 3(r)_{n'3, n2}^2 \}, \\ \sigma_{n3} &= 6 \sum_{n'} (r)_{n'2, n3}^2. \end{aligned} \right\} \quad (3.12)$$

The range of n' in these sums depends upon the value of l' (either 0, 1, 2, or 3) associated with n' , but it includes every occupied subgroup having the quantum numbers n', l' .

The terms $(R^2)_\rho$ are treated similarly, and we find

$$(R^2)_{nlms} = (r^2)_{nl, nl} - \frac{1}{2(2l+1)} \tau_{nl}, \quad (3.13)$$

where τ_{nl} is simply related to σ_{nl} , as it is only necessary to replace $(r)_{n'l, nl}^2$ where it occurs in σ_{nl} by the product $(r^2)_{n'l, nl} (r)_{n'l, nl}$. For example

$$\tau_{n1} = 2 \sum_{n'} \{ (r^2)_{n'0, n1} (r)_{n'0, n1} + 2(r^2)_{n'2, n1} (r)_{n'2, n1} \}.$$

When the atom contains incomplete subgroups, it is necessary first to consider whether the electronic configuration is degenerate and, if so, whether the previous argument needs to be modified. However, if an s shell is incomplete, as in the alkali atoms, no modification is required other than that the contribution of that shell to the σ and τ terms, as defined by (3.12), shall be halved.

The self-consistent atomic fields calculated by Hartree and others provide suitable functions $P(nl|r)$ from which to evaluate $(r^2)_{nl, nl}$, $(r)_{n'l, nl}$, etc., by numerical integration, and we can therefore calculate the polarizability of any atom or ion for which the self-consistent field is known, provided proper care is taken when there is an incomplete subgroup. It is true that the Hartree functions for the same l but different n are not exactly orthogonal,

but the error is found to be too small to make it necessary to derive an orthogonal set. A more important disadvantage is that they neglect electron exchange entirely, and although the self-consistent field equations have been modified by Fock to include exchange, Fock wave functions have as yet been calculated for very few atoms. A comparison of Hartree and Fock functions is made below.

3b—Numerical Results and Discussion

In § 3, three different expressions for the polarizability of an atom were obtained by progressive application of the variation method. We shall now discuss the following points which arise in comparing the polarizabilities calculated from these formulae with each other and with experimental values: the effect of exchange on the calculated polarizabilities; the effect of introducing exchange in the electronic wave functions; the effect of the more accurate variation function leading to (3.9); and Kirkwood's relation between polarizability and magnetic susceptibility. These points will be illustrated with the help of available wave functions for atoms and ions with complete subgroups.

It was shown above that the functions $(R^2)_\rho$ occurring in (3.5) and (3.7) depend only on the quantum numbers n and l and hence the summation within each (n, l) subgroup can be carried out at once. If this is done and (3.11) substituted for $(R^2)_\rho$, then (3.5) and (3.7) rewritten in c.g.s. units are respectively

$$\alpha = \frac{4a_0^3}{9N} \left[\sum_{nl} \{ \nu_l(r^2)_{nl, nl} - \sigma_{nl} \} \right]^2, \quad \text{I}$$

$$\alpha = \frac{4a_0^3}{9} \sum_l \nu_l \{ \nu_l(r^2)_{nl, nl} - \sigma_{nl} \}^2, \quad \text{II}$$

where a_0 is the radius of the first Bohr orbit in the hydrogen atom ($4a_0^3/9 = 0.654 \times 10^{-23}$ cm.³); the terms $(r^2)_{nl, nl}$ and σ_{nl} are still expressed in atomic units. The summations are for all occupied subgroups, and

$$\nu_l = 2(2l+1).$$

Of these results, I was derived from one parameter only, and should be compared with that found by Kirkwood, using for the atomic wave function the simple product

$$\Psi = \psi_{s1} \psi_{p2} \dots \psi_{pN},$$

which takes no account of electron exchange; likewise with II we may compare the result obtained by Hellmann, also neglecting exchange but

using a different parameter for each subgroup of electrons. The formulae of Kirkwood and Hellmann, which we will denote by I A, II A respectively, are

$$\alpha = \frac{4a_0^3}{9N} \left\{ \sum_{nl} \nu_l(r^2)_{nl, nl} \right\}^2, \quad \text{I A}$$

$$\alpha = \frac{4a_0^3}{9} \sum_{nl} \nu_l(r^2)_{nl, nl}^2, \quad \text{II A}$$

in the same units and notation. It will be seen that the only difference is in the absence of the terms σ_{nl}

These four expressions for α exemplify two general principles; first, that the inclusion of electron exchange, which is similar to the removal of a constraint on the system, decreases the energy of perturbation and therefore the polarizability. Thus the values obtained from I and II are respectively less than those obtained from I A and II A, since σ_{nl} is always positive. Secondly, it is generally true in applications of the variation method with a given form for the perturbed wave function that, as more parameters are introduced, the calculated energy approaches the exact value as a lower limit. The energy of polarization being negative, we should expect the calculated polarizability to increase in magnitude as the number of parameters is increased, and, in fact, the values given by II and II A are respectively greater than those given by I and I A. To illustrate these remarks, the known Hartree† fields of Ne, A, Rb⁺ and Cs⁺ have been applied, as suggested in § 3a, to calculate values of α , which are shown in Table I.

TABLE I—CALCULATED ATOMIC POLARIZABILITIES. $\alpha \times 10^{24}$

Method		Ne	A	Rb ⁺	Cs ⁺
I	With exchange	0.348	1.49	1.09	1.43
I A	Without exchange				
One parameter		0.774	3.38	2.54	4.32
II	With exchange	0.517	4.09	4.35	8.44
II A	Without exchange				
Parameter for each subgroup		0.969	7.16	7.47	16.9

Table I shows that the inclusion of exchange terms in the formula for α may reduce the calculated value by as much as 50 %. Before making any comparison with experiment we must also consider the effect of exchange on the electron wave functions, as shown by the difference between Hartree

† I am indebted to Dr J. McDougall for the use of unpublished wave functions for Ne. Other Hartree fields have been used as follows: A, a field calculated by J. McDougall and R. A. Buckingham; Cs⁺ (and also K⁺) from D. R. Hartree (1934); Rb⁺, analytical wave functions derived from the Hartree function by J. C. Slater (1932).

and Fock functions. The few Fock functions which are known indicate the change to be expected; thus D. R. and W. Hartree (1935)[†] have already pointed out that in normal Be the term $(r^2)_{20,20}$ is about 12 % less, and in Na^+ $(r^2)_{21,21}$ is about 13 % less for the Fock field than for the Hartree field. In Cl^- , the decrease in $(r^2)_{21,21}$ is as much as 30 %. The effect of exchange on the σ_{nl} terms can be examined in Na^+ , Cl^- , and it is found that these terms are practically the same for the Fock and Hartree fields. The corresponding values of $\alpha \cdot 10^{24}$ calculated from II are as follows.

	Hartree	Fock
Be	12.0	9.28
Na^+	0.183	0.134
Cl^-	16.2	6.9

The use of Fock functions thus leads to a further decrease in the calculated polarizabilities, and therefore to better agreement with experiment, since Hartree functions are found to give values which are too high. There is, however, one important exception to this general statement: the Hartree field for Na^+ gives $\alpha = 0.183 \times 10^{-24}$, which is slightly larger than the 0.17×10^{-24} deduced by Mayer (1933) from spectroscopic data, but less than the 0.245×10^{-24} given by Heydweiller's (1925) measurements of the refractivities of alkali halides in solution. We are therefore led to consider the more accurate formula (3.9) which, when written in c.g.s. units, is

$$\alpha = \frac{4a_0^3}{9} \sum_{nl} \nu_l \{ \nu_l (r^2)_{nl,nl} - \sigma_{nl} \}^2 (1 + \Delta_{nl}). \quad \text{III}$$

This represents the closest approximation to the actual atomic system which has been considered, as the assumed system satisfies the Pauli Principle and allows two independent parameters for each subgroup. The difficulty in applying it arises from the calculation of the Δ_{nl} terms for which accurate wave functions are necessary.

Nevertheless, the available functions indicate the relative importance of these terms in different atoms. In Table II are given the values of $1 + \Delta_{nl}$, those in the first column being derived by numerical integration from self-consistent functions (Hartree functions, with the exception of Be and Na^+ , for which Fock functions were used), and those in the second column from analytical functions (Slater 1930) of the type $r^{n^*-1} \exp(-n^*r/Z - S)$, in which n^* is an effective quantum number and S a screening constant. Since the inner groups of electrons make a very small contribution to α ,

[†] The Fock field for Na^+ was calculated by V. Fock and M. Petrashen (1934). I am indebted to Professor Hartree for information about the Fock field of Cl^- before it was published (Hartree, D. R. and W. 1936).

only the outer group in each atom has been considered, although a calculation for Na^+ shows that the contribution of an inner S -group may be doubled. The large discrepancy between the two values for the $2p$ group of neon may be understood by comparing the theoretical values of the magnetic susceptibility χ given by the same wave functions: thus the Slater function underestimates the extension of the $2p$ charge distribution from the nucleus and gives a susceptibility less than that observed, whereas the Hartree function does the opposite. Hence the proper value of $1 + \Delta(2p)$ probably lies between those given. For Na^+ , however, both the Slater and Fock functions† appear to underestimate χ , so that the correct value of $1 + \Delta(2p)$ may be greater than 1.15. But the general conclusion is that the value of $1 + \Delta_{nl}$ for an outer group differs appreciably from unity only for the lighter atoms such as He, Ne, and Na^+ . Its effect is then to increase the calculated polarizability, and for Na^+ the agreement with experiment is thereby slightly improved.

TABLE II

		Self-consistent functions	Slater
He	$1 + \Delta(1s)$	—	1.125
Be	$1 + \Delta(2s)$	1.00	1.01
Ne	$1 + \Delta(2p)$	1.40	1.03
A	$1 + \Delta(3p)$	1.00 _h	1.00
Kr	$1 + \Delta(4p)$	—	1.00 _h
Xe	$1 + \Delta(5p)$	—	1.01
Na^+	$1 + \Delta(2s)$	2.03	2.83
	$1 + \Delta(2p)$	1.15	1.03
K^+	$1 + \Delta(3p)$	1.03	1.00

It is perhaps worth mentioning an empirical method of estimating these correction terms. It was pointed out by Kirkwood that polarizability and magnetic susceptibility are related quantities and using his result for α (formula I A) he suggested the relation

$$-\chi = \frac{Le^2 a_0^3}{4mc^2} \sqrt{N\alpha}.$$

Further developments by the variation method show that this is hardly justified, though as an empirical relation it is more successful than one would expect (Brindley 1933). However, we may make use of the result that the main part of α comes from an outer group of electrons, say ν_0 in

† The Slater and Fock functions give $\chi = -4.17, 5.0 \times 10^{-8}$ respectively. The experimental value for ions in solution is about -5.7×10^{-8} , and for crystal ions about -6.1×10^{-8} .

number; then if $(r^2)_0$ and σ_0 are the values of $(r^2)_{nl, nl}$ and σ_{nl} for this group, its contribution to χ is

$$\chi_0 = -\frac{Le^2 a_0^2}{6mc^2} \nu_0 (r^2)_0. \quad (3.14)$$

By comparing (3.14) and III, we find that approximately

$$\left(\frac{9\nu_0 \alpha}{4a_0^2 (1 + \Delta'_0)} \right)^{\frac{1}{2}} + \sigma_0 = -\frac{6mc^2}{Le^2 a_0^2} \chi_0. \quad (3.15)$$

Now suppose α and χ are known experimentally, and that χ_0 can be estimated from χ , as when the relative importance of the different electron groups is known from the self-consistent field. If in addition σ_0 , which is usually much smaller than the other terms, is given its theoretical value, then the relation (3.15) can be applied to find $1 + \Delta'_0$. The results in Table III have been derived in this way.

TABLE III†

	$\alpha \times 10^{24}$	$-\chi \times 10^4$	$-\chi_0 \times 10^4$	σ_0	$1 + \Delta'_0$
Ne	0.392	6.75	5.37	1.79	1.43
Na ⁺	0.280	6.1	4.56	1.23	1.23
Cl ⁻	3.01	24.2	18.9	5.74	0.84
A	1.63	19.54	14.7	4.77	0.79
K ⁺	1.09	14.6	10.7	3.67	1.03
K ⁻	2.46 ₅	28.0	17.4	6.5‡	0.95
Cs ⁺	2.72	35.1	19.3	8.37	0.96

† For the rare gases, the experimental values of α are taken from C. and M. Cuthbertson (1911) and of χ from Mann (1936). The data for the ions refer to ions in the crystalline state; the values of χ are those given by Brindley and Hoare (1935), and the values of α are derived from the refractivities of the alkali halide crystals, measured by Spangenberg (1923).

‡ The value of σ_0 for Kr is calculated from approximate self-consistent wave functions for the 4s and 4p electrons, kindly sent to me by Professor D. R. Hartree.

Direct comparison of the values of $1 + \Delta'_0$ with the theoretical values $1 + \Delta_0$ is hardly justified, since the error of approximation involved in (3.15) is not known. Further errors may be introduced by inaccuracy in the wave functions, affecting σ_0 and the ratio χ_0/χ , and in the experimental data, which for the ions are especially uncertain. But although the values in Table III must be accepted with caution, they do confirm the theoretical result that the Δ terms are important for the lighter atoms Ne and Na⁺. It is, however, rather surprising that the Δ'_0 value for argon should be so large and negative.

APPENDIX

Much of the integration required in § 3 is applicable to other problems, and so we shall give a separate account of the methods used, beginning with the integrals, such as those in (3.2), which contain the density function $\Psi^*\Psi$. This is replaced by the determinant A obtained by multiplication of the determinants Ψ^* and Ψ , defined by (3.1). A has the form

$$A = \begin{vmatrix} a_{11} a_{12} & \dots & a_{1N} \\ a_{21} a_{22} & \dots & a_{2N} \\ \vdots & & \vdots \\ a_{N1} a_{N2} & \dots & a_{NN} \end{vmatrix}, \quad (1)$$

where $a_{pq} = \sum_p \psi_p^* \psi_{pq}$, the summation including all occupied states from α to ν , and A has the property that it can easily be integrated over the co-ordinate space of any electron (see Lennard-Jones 1930*b*). For instance, if $f(p)$ is a function of the co-ordinates of the p th electron only, then by completing the integration for all electrons except the p th, it is easily proved that

$$\frac{1}{N!} \int \Psi^* f(p) \Psi d\tau = \frac{1}{N} \int f(p) a_{pp} d\tau_p,$$

a result which is independent of the choice of p . We can therefore drop the electron suffix, and remove the factor $1/N$ by summing for p from 1 to N ; thus

$$\frac{1}{N!} \sum_p \int \Psi^* f(p) \Psi d\tau = \sum_p \int \psi_p^* f \psi_p d\tau. \quad (2)$$

Similarly if $g(p)$, $g'(q)$ are respectively functions of the co-ordinates of the p th and q th electrons only, then by integrating for all electrons except the p th and q th

$$\begin{aligned} \frac{1}{N!} \int \Psi^* g(p) g'(q) \Psi d\tau &= \frac{1}{N(N-1)} \iint g(p) g'(q) \left| \begin{smallmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{smallmatrix} \right| d\tau_p d\tau_q \\ &= \frac{1}{N(N-1)} \sum_p \sum_q \iint g(p) g'(q) \{ \psi_p^* \psi_{pp} \psi_{pq}^* \psi_{qq} \\ &\quad - \psi_{pq}^* \psi_{pq} \psi_{qp}^* \psi_{qp} \} d\tau_p d\tau_q. \end{aligned}$$

This is likewise independent of the choice of p and q , and hence, dropping the electron suffices,

$$\begin{aligned} \frac{1}{N!} \sum_p \sum_{q \neq p} \int \Psi^* g(p) g'(q) \Psi d\tau \\ = \sum_p \sum_{\sigma} \left[\int \psi_p^* g \psi_p d\tau \int \psi_{\sigma}^* g' \psi_{\sigma} d\tau' - \int \psi_p^* g \psi_{\sigma} d\tau \int \psi_{\sigma}^* g' \psi_p d\tau' \right]. \quad (3) \end{aligned}$$

Turning now to (3.2), we require the integral of v^2 where $v = -\sum_p (\mathbf{r}_p \cdot \mathbf{F})$. v^2 may be written

$$\sum_p \{(\mathbf{r}_p \cdot \mathbf{F})^2 + \sum_{q \neq p} (\mathbf{r}_p \cdot \mathbf{F})(\mathbf{r}_q \cdot \mathbf{F})\}$$

and so from (2) and (3) we get at once

$$\begin{aligned} \frac{1}{N!} \int \Psi^* v^2 \Psi d\tau = \sum_p \left[\int (\mathbf{r} \cdot \mathbf{F})^2 \psi_p^* \psi_p d\tau \right. \\ \left. + \sum_{\sigma \neq p} \left\{ \int (\mathbf{r} \cdot \mathbf{F}) \psi_p^* \psi_p d\tau \int (\mathbf{r}' \cdot \mathbf{F}) \psi_{\sigma}^* \psi_{\sigma} d\tau' \right. \right. \\ \left. \left. - \int (\mathbf{r} \cdot \mathbf{F}) \psi_p^* \psi_{\sigma} d\tau \int (\mathbf{r}' \cdot \mathbf{F}) \psi_{\sigma}^* \psi_p d\tau' \right\} \right]. \end{aligned}$$

Let us introduce a system of rectangular axes in which \mathbf{r} has direction cosines l, m, n , and \mathbf{F} has direction-cosines α, β, γ . Then if r, F are the magnitudes of \mathbf{r} and \mathbf{F}

$$\int \psi_p^* (\mathbf{r} \cdot \mathbf{F})^2 \psi_p d\tau = F^2 \int \psi_p^* r^2 (l\alpha + m\beta + n\gamma)^2 \psi_p d\tau,$$

and when we average for all values of l, m, n , i.e. for all orientations of the atom relative to the direction of \mathbf{F} , this integral reduces to

$$\frac{1}{3} F^2 (r^2)_{pp}.$$

Now $(\mathbf{r} \cdot \mathbf{F})$ is an odd function of r and so its average value with respect to the charge density $\psi_p^* \psi_p$ is zero; the second term in (4) therefore vanishes. The remaining integrals can be treated like the first, and finally

$$\frac{1}{N!} \int \Psi^* v^2 \Psi d\tau = \frac{1}{3} F^2 \sum_p [(r^2)_{pp} - \sum_{\sigma \neq p} \{(x)_{\sigma p} (x)_{p\sigma} + (y)_{\sigma p} (y)_{p\sigma} + (z)_{\sigma p} (z)_{p\sigma}\}].$$

We also require the integral of $(\text{grad}_p v^2)$, in which $\text{grad}_p v = -\text{grad}_p (\mathbf{r}_p \cdot \mathbf{F})$. From (2) it follows that

$$\frac{1}{N!} \sum_p \int \Psi^* (\text{grad}_p v^2) \Psi d\tau = \sum_p \int \psi_p^* \{(\text{grad}(\mathbf{r} \cdot \mathbf{F}))^2\} \psi_p d\tau,$$

and the right-hand side is easily shown to be equal to N . These results lead at once to equation (3.4).

When it is necessary to use equation (2.5) for the perturbation energy, rather more complicated integrals occur but they can be reduced in essentially the same way. Instead of the charge density $\Psi^*\Psi$ we usually have $\psi_{\rho q}^* D_{\rho q}^* D_{\rho p} \psi_{\rho p}$, in which $D_{\rho p}$ is the first minor of $\psi_{\rho p}$ in the determinant Ψ . $D_{\rho q}^*$ and $D_{\rho p}$ are therefore determinants with $(N-1)$ rows and columns and their product is likewise a determinant $(N-1)$ rows and columns, having the form

$$D_{\rho q}^* D_{\rho p} = (-)^{p+q} \begin{vmatrix} a'_{11} & \dots & a'_{1,p-1} & a'_{1,p+1} & \dots & a'_{1N} \\ \vdots & & & & & \vdots \\ a'_{q-1,1} & & & & & a'_{q-1,N} \\ a'_{q+1,1} & & & & & a'_{q+1,N} \\ \vdots & & & & & \vdots \\ a'_{N1} & \dots & a'_{N,p-1} & a'_{N,p+1} & \dots & a'_{NN} \end{vmatrix},$$

where $a'_{pq} = \sum_{\sigma \neq \rho} \psi_{\sigma p}^* \psi_{\sigma q}$, the summation including all states from α to ν except ρ . The product of this determinant with $\psi_{\rho q}^* \psi_{\rho q}$ is, like (1), easily integrated over the co-ordinate space of any electron, and we obtain the following useful results.

$$\left. \begin{aligned} \sum_p \int \psi_{\rho q}^* D_{\rho q}^* D_{\rho p} \psi_{\rho p} f(p) d\tau &= 0, \text{ when } q \neq p, \\ &= N! \int \psi_{\rho}^* f \psi_{\rho} d\tau, \text{ when } q = p, \\ \sum_p \sum_{q \neq p} \int \psi_{\rho q}^* D_{\rho q}^* D_{\rho p} \psi_{\rho p} g(p) g'(t) d\tau &= 0, \text{ when } t \neq q, \\ &= N! \sum_{\sigma \neq \rho} \left\{ \int \psi_{\rho}^* g \psi_{\rho} d\tau \int \psi_{\sigma}^* g' \psi_{\sigma} d\tau - \int \psi_{\rho}^* g \psi_{\sigma} d\tau \int \psi_{\sigma}^* g' \psi_{\rho} d\tau \right\}, \\ &\quad \text{when } t = q. \end{aligned} \right\} \quad (4)$$

Integrals which involve $\psi_{\rho q}^* D_{\rho q}^* D_{\sigma p} \psi_{\sigma p}$ ($\sigma \neq \rho$) can also be reduced by the same method, and we find that

$$\left. \begin{aligned} \int \psi_{\rho p}^* D_{\rho p}^* D_{\sigma p} \psi_{\sigma p} d\tau &= 0, \\ \int \psi_{\rho q}^* D_{\rho q}^* D_{\sigma p} \psi_{\sigma p} f(p) d\tau &= 0, \\ \int \psi_{\rho q}^* D_{\rho q}^* D_{\sigma p} \psi_{\sigma p} g(p) g'(q) d\tau &= 0. \end{aligned} \right\} \quad (5)$$

Let us now consider the integral of $\Psi^* v \Phi$, where $v = \sum_p v_p$ as above, and $\Phi = \sum_p \lambda_p \sum_p \psi_{pp} D_{pp} v_p$. We write

$$\Psi^* v \Phi = \Psi^* (\sum_i v_i) \sum_p \lambda_p \sum_p \psi_{pp} D_{pp} v_p,$$

in which the coefficient of λ_p is

$$\Psi^* \sum_p \psi_{pp} D_{pp} v_p (\sum_i v_i).$$

If Ψ^* in this expression is replaced by its expansion in terms of the elements of the ρ th row, i.e. by $\sum_q \psi_{\rho q}^* D_{\rho q}^*$, then

$$\Psi^* v \Phi = \sum_p \lambda_p \{ \sum_q \sum_p \psi_{\rho q}^* D_{\rho q}^* D_{pp} \psi_{pp} (v_p^2 + \sum_{i \neq p} v_p v_i) \}.$$

When we apply equations (4) and insert $v_p = -(\mathbf{r}_p \cdot \mathbf{F})$ the terms which do not vanish on integration give

$$\frac{1}{N!} \int \Psi^* v \Phi d\tau = \sum_p \lambda_p \left[\int \psi_{\rho}^* (\mathbf{r} \cdot \mathbf{F})^2 \psi_{\rho} d\tau - \sum_{\sigma \neq \rho} \int \psi_{\rho}^* (\mathbf{r} \cdot \mathbf{F}) \psi_{\sigma} d\tau \int \psi_{\sigma}^* (\mathbf{r}' \cdot \mathbf{F}) \psi_{\rho} d\tau' \right],$$

which by our previous results

$$= \frac{1}{2} F^2 \sum_p \lambda_p (R^2)_p.$$

From (2.5) we have also the following term involving grad_p :

$$\int \text{grad}_p \Phi^* \cdot \text{grad}_p \Phi d\tau - \frac{1}{N!} \int \Phi^* \Phi d\tau \int \text{grad}_p \Psi^* \cdot \text{grad}_p \Psi d\tau.$$

Since $\Phi = \sum_p \lambda_p \sum_p \psi_{pp} D_{pp} v_p$ and D_{pp} contains the co-ordinates of all electrons except the p th,

$$\text{grad}_p \Phi = \sum_p \lambda_p \{ D_{pp} (\psi_{pp} \text{grad}_p v_p + v_p \text{grad}_p \psi_{pp}) + \sum_{q \neq p} \psi_{\rho q} v_q \text{grad}_p D_{\rho q} \}.$$

On integration, we neglect all terms except that in $(\text{grad}_p v_p)^2$ assuming that the remainder, which with the integral of $\text{grad}_p \Psi^* \cdot \text{grad}_p \Psi$ are analogous to (2.6), are small. Approximately therefore

$$\begin{aligned} \frac{1}{N!} \int \left\{ \int \text{grad}_p \Phi^* \cdot \text{grad}_p \Phi d\tau - \frac{1}{N!} \int \Phi^* \Phi d\tau \int \text{grad}_p \Psi^* \cdot \text{grad}_p \Psi d\tau \right\} \\ = \frac{1}{N!} \sum_p \int \sum_p \lambda_p D_{pp}^* \psi_{pp}^* (\sum_{\sigma} \lambda_{\sigma} D_{\sigma p} \psi_{\sigma p}) (\text{grad}_p v_p)^2 d\tau \\ = \frac{1}{N!} \sum_p \sum_p \lambda_p^2 \int \psi_{pp}^* D_{pp}^* D_{pp} \psi_{pp} (\text{grad}_p v_p)^2 d\tau \quad \text{by (5)} \\ = \sum_p \lambda_p^2 \int \psi_{\rho}^* (\text{grad}(\mathbf{r} \cdot \mathbf{F}))^2 \psi_{\rho} d\tau \quad \text{by (4)} \\ = F^2 \sum_p \lambda_p^2. \end{aligned}$$

Equation (3.6) for the energy of polarization follows at once from these results.

The integrals which arise when we introduce the more accurate perturbed functions described in § 3 do not require any extension of the method outlined above, and we shall therefore not give further details.

I am very grateful to Professor J. E. Lennard-Jones, F.R.S., for many helpful discussions, also to the Department of Scientific and Industrial Research for a grant during a period when many of these calculations were made.

SUMMARY

The theory of the polarization of an atom by a uniform electric field is developed, using a variation method introduced by Kirkwood, and the atomic polarizabilities of several atoms containing closed electron groups calculated from their self-consistent fields. It is verified that an anti-symmetrical wave function for the atom gives better results than a symmetrical function. Higher approximations, in which more parameters are introduced in the perturbed wave functions, are considered. Calculations with the Fock fields of the atoms of Be, Na⁺, and Cl⁻ show that the inclusion of exchange in the electron wave functions has considerable effect on the calculated polarizabilities.

REFERENCES

- Atanasoff 1930 *Phys. Rev.* **36**, 1232.
 Brindley, G. W. 1933 *Phys. Rev. A*, **43**, 1030.
 Brindley, G. W. and Hoare, F. E. 1935 *Proc. Roy. Soc. A*, **152**, 342.
 Cuthbertson, C. and M. 1911 *Proc. Roy. Soc. A*, **84**, 13.
 Eisenschütz, R. and London, F. 1930 *Z. Phys.* **60**, 491.
 Fock, V. and Petrashen, M. 1934 *Phys. Z. Sowjet.* **6**, 368.
 Hartree, D. R. 1934 *Proc. Roy. Soc. A*, **143**, 508.
 Hartree, D. R. and W. 1935 *Proc. Roy. Soc. A*, **150**, 9.
 — — 1936 *Proc. Roy. Soc. A*, **156**, 45.
 Hassé, H. R. 1930 *Proc. Camb. Phil. Soc.* **26**, 542.
 — — 1931 *Proc. Camb. Phil. Soc.* **27**, 66.
 Hellmann, H. 1935 *Acta Physicochim. U.S.S.R.* **2**, 273.
 Heydweiller 1925 *Phys. Z.* **26**, 526.
 Kirkwood, J. G. 1932 *Phys. Z.* **33**, 57.
 Lennard-Jones, J. E. 1930a *Proc. Roy. Soc. A*, **129**, 598.
 — — 1930b *Proc. Camb. Phil. Soc.* **27**, 469.
 London, F. 1930 *Z. phys. Chem. B*, **11**, 222.
 Mann, K. G. 1936 *Z. Phys.* **98**, 548.
 Margenau, H. 1931 *Phys. Rev.* **38**, 747.
 Mayer, J. E. 1933 *J. Chem. Phys.* **1**, 270.

Pauling, L. and Beach, J. Y. 1935 *Phys. Rev.* **47**, 686.

Slater, J. C. 1930 *Phys. Rev.* **36**, 57.

— 1932 *Phys. Rev.* **42**, 33.

Slater, J. C. and Kirkwood, J. G. 1931 *Phys. Rev.* **37**, 686.

Spangenberg, K. 1923 *Z. Kristallogr.* **57**, 517.

Vinti, P. 1932 *Phys. Rev.* **41**, 813.

Wang, S. C. 1927 *Phys. Z.* **28**, 663.

The Quantum Theory of Atomic Polarization

II—The van der Waals Energy of Two Atoms

BY R. A. BUCKINGHAM, B.A., *Queen's University, Belfast*

(Communicated by J. E. Lennard-Jones, F.R.S.—Received 29 December 1936)

In a previous paper (p. 94) (which will be referred to as Paper I), the polarizability of an atom in a uniform electric field was calculated by a method of varying parameters. The same method can equally well be applied to find the energy of interaction of two atoms a large distance apart, by treating their interaction as a perturbation of the system in which the atoms are separated by an infinite distance. The mutual energy, other than that arising from the ionic charges, if they exist, of the atoms, is usually called the van der Waals energy.

We shall suppose that one atom contains N_1 electrons, and is represented in its unperturbed state by a determinantal wave function Ψ_1 (as in equation (3.1) of Paper I), containing electronic functions of the type $\psi_{\rho_1 p}$, where ρ_1 may be any one of the N_1 occupied states $\alpha_1, \beta_1, \dots, \nu_1$ of this atom, and p denotes that the spatial and spin co-ordinates of the p th electron are inserted in $\psi_{\rho_1 p}$. Similarly the second atom contains N_2 electrons, and is represented in its unperturbed state by a determinant Ψ_2 , containing functions $\psi_{\rho_2 r}$, where ρ_2 may be one of the N_2 occupied states $\alpha_2, \beta_2, \dots, \nu_2$, and r refers to the co-ordinates of the r th electron. Then if we neglect any exchange of electrons between the atoms, the wave function of the unperturbed system of two atoms is

$$\Psi = \Psi_1 \Psi_2$$

and

$$\int \Psi^* \Psi d\tau = N_1! N_2!,$$

where the integration is taken over the co-ordinate-space of all N_1 electrons of atom 1 and all N_2 electrons of atom 2.

When the atoms interact at a large distance R , it is convenient to split the total interaction into a number of components, one such component being the interaction of the p th electron in atom 1 and unit positive charge at the nucleus of atom 1 with the r th electron in atom 2 and unit charge at the nucleus of atom 2. Let the resulting increase of the potential field acting on the p th electron in atom 1 be a small amount v_{pr} . Then the total perturbing field acting on this electron is

$$\sum_{r=1}^{N_2} v_{pr},$$

and similarly the perturbing field acting on the r th electron in atom 2 is

$$\sum_{p=1}^{N_1} v_{pr} \quad (v_{pr} = v_{rp}).$$

If we adopt the same crude first approximation of Paper I (§3), the perturbed wave functions of the atoms are obtained by replacing

$$\psi_{p,1} \text{ by } \psi_{p,1}(1 + \lambda_1 \sum_r v_{pr}) \text{ in } \Psi_1$$

and

$$\psi_{r,2} \text{ by } \psi_{r,2}(1 + \lambda_2 \sum_p v_{pr}) \text{ in } \Psi_2,$$

where λ_1, λ_2 are parameters, and the perturbed wave function of the system, when only linear terms in v_{pr} are retained, is

$$\Psi(1 + \lambda v), \quad (1)$$

where $\lambda = \lambda_1 + \lambda_2$ and $v = \sum_{p=1}^{N_1} \sum_{r=1}^{N_2} v_{pr}$.

When R is large, v_{pr} may be expanded as a series of ascending powers of $1/R$. If \mathbf{t} is a unit vector along the line of the atomic nuclei, \mathbf{r}_p the position vector of the p th electron in atom 1 referred to nucleus 1, and \mathbf{r}_r the position vector of the r th electron in atom 2 referred to nucleus 2, then in atomic units

$$v_{pr} = \frac{1}{R^3} \{ \mathbf{r}_p \cdot \mathbf{r}_r - 3 \mathbf{r}_p \cdot \mathbf{t} \mathbf{r}_r \cdot \mathbf{t} \} \\ + \frac{3}{2R^4} \{ r_p^2 \mathbf{r}_r \cdot \mathbf{t} - r_r^2 \mathbf{r}_p \cdot \mathbf{t} + (\mathbf{r}_p \cdot \mathbf{t} - \mathbf{r}_r \cdot \mathbf{t}) (2 \mathbf{r}_p \cdot \mathbf{r}_r - 5 \mathbf{r}_p \cdot \mathbf{t} \mathbf{r}_r \cdot \mathbf{t}) \} + O\left(\frac{1}{R^5}\right). \quad (2)$$

The term in R^{-3} is usually referred to as the dipole-dipole interaction and that in R^{-4} as the dipole-quadrupole interaction. We shall proceed to consider each term separately.

1—DIPOLE-DIPOLE ENERGY

The van der Waals energy is given by a modified form of equation (2.7) of Paper I, viz.:

$$\bar{h} = \frac{1}{N_1! N_2!} \int \Psi^* \left[2\lambda v^2 + \frac{1}{2}\lambda^2 \left(\sum_{p=1}^{N_1} (\text{grad}_p v)^2 + \sum_{r=1}^{N_2} (\text{grad}_r v)^2 \right) \right] \Psi d\tau. \quad (3)$$

With v as defined above, and the dipole-dipole term for v_{pr} , (3) is reducible to a sum of simple integrals† and can be averaged for random orientation of each atom with respect to the direction of \mathbf{t} . With the same notation, i.e.

$$(f)_{\rho\rho} = \int \psi_\rho^* f \psi_\rho d\tau, \quad (g)_{\rho\rho'} = \int \psi_\rho^* g \psi_{\rho'} d\tau,$$

we obtain

$$\left. \begin{aligned} \frac{1}{N_1! N_2!} \int \Psi^* v^2 \Psi d\tau &= \frac{2}{3R^6} \bar{R}_1^2 \bar{R}_2^2, \\ \frac{1}{N_1! N_2!} \sum_p \int \Psi^* (\text{grad}_p v)^2 \Psi d\tau &= \frac{2}{R^6} N_1 \bar{R}_2^2, \\ \frac{1}{N_1! N_2!} \sum_r \int \Psi^* (\text{grad}_r v)^2 \Psi d\tau &= \frac{2}{R^6} N_2 \bar{R}_1^2, \end{aligned} \right\} \quad (4)$$

$$\bar{R}_1^2 = \sum_{\rho_1} (R^2)_{\rho_1}, \quad \bar{R}_2^2 = \sum_{\rho_2} (R^2)_{\rho_2},$$

where

$$(R^2)_{\rho_1} = (r^2)_{\rho_1 \rho_1} - \sum_{\sigma_1 \neq \rho_1} \{ (x)_{\sigma_1 \rho_1} (x)_{\rho_1 \sigma_1} + (y)_{\sigma_1 \rho_1} (y)_{\rho_1 \sigma_1} + (z)_{\sigma_1 \rho_1} (z)_{\rho_1 \sigma_1} \},$$

$$(R^2)_{\rho_2} = (r^2)_{\rho_2 \rho_2} - \sum_{\sigma_2 \neq \rho_2} \{ (x)_{\sigma_2 \rho_2} (x)_{\rho_2 \sigma_2} + (y)_{\sigma_2 \rho_2} (y)_{\rho_2 \sigma_2} + (z)_{\sigma_2 \rho_2} (z)_{\rho_2 \sigma_2} \}.$$

Combining these results, equation (3) for \bar{h} becomes

$$\bar{h} = \frac{2}{R^6} \left\{ \frac{1}{3} \lambda \bar{R}_1^2 \bar{R}_2^2 + \frac{1}{2} \lambda^2 (N_2 \bar{R}_1^2 + N_1 \bar{R}_2^2) \right\}. \quad (5)$$

The condition that \bar{h} shall be a minimum, viz. $d\bar{h}/d\lambda = 0$, gives

$$\lambda = -\frac{1}{3} \frac{\bar{R}_1^2 \bar{R}_2^2}{(N_2 \bar{R}_1^2 + N_1 \bar{R}_2^2)},$$

and

$$\bar{h}_{\min} = -\frac{4}{9R^6} \frac{(\bar{R}_1^2)^2 (\bar{R}_2^2)^2}{N_2 \bar{R}_1^2 + N_1 \bar{R}_2^2},$$

or, if we define c_{12} by $\bar{h}_{\min} = -c_{12}/R^6$, then

$$c_{12} = \frac{4}{9} \frac{(\bar{R}_1^2)^2 (\bar{R}_2^2)^2}{N_2 \bar{R}_1^2 + N_1 \bar{R}_2^2}, \quad (6)$$

in atomic units.

† An account of the method is contained in an Appendix to Paper I.

There follows at once a simple relation between the van der Waals constant c_{12} and the polarizabilities α_1, α_2 of the two atoms, since by equation (3.5) of Paper I

$$\alpha_1 = \frac{4}{9N_1} (\bar{R}_1^2)^2 \quad \text{and} \quad \alpha_2 = \frac{4}{9N_2} (\bar{R}_2^2)^2$$

and therefore

$$c_{12} = \frac{3}{2} \frac{\alpha_1 \alpha_2}{\sqrt{\frac{\alpha_1}{N_1}} + \sqrt{\frac{\alpha_2}{N_2}}} \quad (7)$$

This relation was also obtained by Kirkwood (1932) neglecting exchange effects.

The next stage is to extend the analysis of the van der Waals energy of two atoms as was done for the polarizability of one atom, by increasing the number of parameters. We again assume that the wave function of the system of two unperturbed atoms is $\Psi_1 \Psi_2$ and obtain the perturbed function by replacing

$$\psi_{\rho_1 p} \quad \text{by} \quad \psi_{\rho_1 p} (1 + \lambda_{\rho_1} \sum_r v_{pr}) \quad \text{in} \quad \Psi_1$$

and

$$\psi_{\rho_2 r} \quad \text{by} \quad \psi_{\rho_2 r} (1 + \lambda_{\rho_2} \sum_p v_{pr}) \quad \text{in} \quad \Psi_2,$$

where λ_{ρ_1} is a different constant for each state in atom 1 from α_1 to ν_1 and λ_{ρ_2} for each state in atom 2 from α_2 to ν_2 . When quadratic and higher powers of v_{pr} are neglected, the perturbed function becomes

$$\Psi_1 \Psi_2 + \Psi_1 \Phi_2 + \Psi_2 \Phi_1,$$

where

$$\Phi_1 = \sum_{\rho_1} \lambda_{\rho_1} \sum_p \psi_{\rho_1 p} D_{\rho_1 p} (\sum_r v_{pr}), \quad \Phi_2 = \sum_{\rho_2} \lambda_{\rho_2} \sum_r v_{\rho_2 r} D_{\rho_2 r} (\sum_p v_{pr}),$$

p and r being summed from 1 to N_1 and from 1 to N_2 respectively. $D_{\rho_1 p}$ is the first minor of $\psi_{\rho_1 p}$ in Ψ_1 and $D_{\rho_2 r}$ the first minor of $\psi_{\rho_2 r}$ in Ψ_2 .

The equation for the energy which takes the place of (3) is

$$\begin{aligned} \bar{h} = & \frac{1}{N_1! N_2!} \int (\Psi^* \Phi + \Phi^* \Psi) v \, d\tau + \frac{1}{2N_1! N_2!} \sum_p \left[\int \text{grad}_p \Phi^* \cdot \text{grad}_p \Phi \, d\tau \right. \\ & \left. - \frac{1}{N_1! N_2!} \int \Phi^* \Phi \, d\tau \int \text{grad}_p \Psi^* \cdot \text{grad}_p \Psi \, d\tau \right] \\ & + \text{similar terms in grad}_r \end{aligned} \quad (8)$$

in which v is the same as in (3), $\Psi = \Psi_1 \Psi_2$ and $\Phi = \Psi_1 \Phi_2 + \Phi_1 \Psi_2$.

The first integral in this expression reduces to

$$\frac{2}{3} \cdot \frac{2}{R^6} \sum_{\rho_1} \sum_{\rho_2} (\lambda_{\rho_1} + \lambda_{\rho_2}) (R^2)_{\rho_1} (R^2)_{\rho_2},$$

where $(R^2)_{\rho_1}$, $(R^2)_{\rho_2}$ are as defined in (4). The contribution to \bar{h} of the integrals in grad_p is

$$\frac{1}{2} \cdot \frac{2}{R^6} \sum_{\rho_1} \sum_{\rho_2} (\lambda_{\rho_1} + \lambda_{\rho_2})^2 (R^2)_{\rho_2},$$

and that of the integrals in grad_r is similarly

$$\frac{1}{2} \cdot \frac{2}{R^6} \sum_{\rho_1} \sum_{\rho_2} (\lambda_{\rho_1} + \lambda_{\rho_2})^2 (R^2)_{\rho_1}.$$

When these results are substituted in (8),

$$\bar{h} = \frac{2}{R^6} \sum_{\rho_1} \sum_{\rho_2} [\frac{1}{2}(\lambda_{\rho_1} + \lambda_{\rho_2}) (R^2)_{\rho_1} (R^2)_{\rho_2} + \frac{1}{2}(\lambda_{\rho_1} + \lambda_{\rho_2})^2 \{(R^2)_{\rho_1} + (R^2)_{\rho_2}\}] \quad (9)$$

and hence, taking the independent parameters to be $\lambda_{\rho_1} + \lambda_{\rho_2}$ we obtain

$$c_{12} = \frac{4}{9} \sum_{\rho_1} \sum_{\rho_2} \frac{(R^2)_{\rho_1}^2 (R^2)_{\rho_2}^2}{(R^2)_{\rho_1} + (R^2)_{\rho_2}}. \quad (10)$$

A relation between c_{12} and the polarizabilities follows if we know the contributions to the polarizabilities made by each subgroup. Thus let the contribution to α_1 , as given by (3.7) of Paper I of the (n_1, l_1) subgroup in atom 1 be $\alpha_{n_1 l_1}$ and the contribution to α_2 of the (n_2, l_2) subgroup in atom 2 be $\alpha_{n_2 l_2}$ and then

$$c_{12} = \frac{2}{3} \sum_{n_1 l_1} \sum_{n_2 l_2} \frac{\alpha_{n_1 l_1} \alpha_{n_2 l_2}}{\sqrt{\frac{\alpha_{n_1 l_1}}{\nu_{l_1}}} + \sqrt{\frac{\alpha_{n_2 l_2}}{\nu_{l_2}}}} \quad (11)$$

in which all pairs of subgroups are included in the summation. For completely occupied subgroups, $\nu_{l_1} = 2(2l_1 + 1)$ and $\nu_{l_2} = 2(2l_2 + 1)$.

The relation (7), of which (11) is the more general form, corresponds to the case when the electrons in each atom are considered as a single group. One point to notice is that the form of the relation between c_{12} and the α 's is unaffected by the inclusion of electron exchange, and depends only on the number of parameters used in applying the variation method. Moreover, an increase in the number of parameters, generally speaking, lowers the van der Waals constants which one calculates from given polarizabilities, and so smaller constants are obtained from (11) than from (7). Even these smaller constants should be regarded as an upper limit to the correct dipole-dipole constants.

It is possible to carry through the calculation of van der Waals energy with the more accurate form of perturbed wave function† which was intro-

† See Paper I, § 3. The perturbed form of ψ_{pp} is taken to be

$$\psi_{ps}\{1 + \lambda_p v_p(1 + \mu_p r_p)\}.$$

duced previously in calculating polarizabilities. Each term in (10) is consequently multiplied by a factor $1 + \Delta_{\rho_1 \rho_2}$, where $\Delta_{\rho_1 \rho_2}$ is a complicated expression involving the wave functions ρ_1 and ρ_2 , but though the resulting formula is undoubtedly more accurate than either (8) or (10), it is not of much practical value unless very good wave functions are available. It is not possible to deduce a simple relation between c_{12} and the polarizabilities; however, some estimates of the $\Delta_{\rho_1 \rho_2}$ terms using Slater functions suggest that the van der Waals constants given by (11) would not be reduced by more than a few per cent, except possibly for helium and similar light atoms.

We have made use of the observed polarizabilities of the rare-gas atoms and alkali ions to calculate the van der Waals constants from (11), assuming the *relative* contributions to α of the various subgroups to be the same as for the self-consistent field. Moreover, it is unlikely that the relative magnitude of these contributions will differ appreciably in say, Xs and Cs⁺, which have the same electronic structure, and so, although the self-consistent field for Xe is not known, we have deduced the van der Waals constants for Xe with other atoms. The values of c_{12} (in erg cm³) are given in Table I, the change from atomic to c g s. units being achieved by multiplying (4.11) by $e^2 a_0^3$. The results are discussed at the end of the paper.

TABLE I.—THE VAN DER WAALS CONSTANTS, CALCULATED FROM OBSERVED POLARIZABILITIES†

	$c_{12} \times 10^{40}$ erg cm. ³		$c_{12} \times 10^{40}$ erg cm. ³
He—He	1.63	Li ⁺ —Li ⁺	0.097
He—Ne	3.48	Li ⁺ —Na ⁺	0.584
He—A	9.89	Li ⁺ —K ⁺	1.51
He—Kr	14.4	Li ⁺ —Rb ⁺	2.21
He—Xe	20.7	Li ⁺ —Cs ⁺	3.14
Ne—Ne	7.48	Na ⁺ —Na ⁺	3.70
Ne—A	20.5	Na ⁺ —K ⁺	10.4
Ne—Kr	30.0	Na ⁺ —Rb ⁺	15.1
Ne—Xe	42.6	Na ⁺ —Cs ⁺	21.6
A—A	63.5	K ⁺ —K ⁺	33.3
A—Kr	92.7	K ⁺ —Rb ⁺	48.4
A—Xe	135.5	K ⁺ —Cs ⁺	71.3
Kr—Kr	136	Rb ⁺ —Rb ⁺	70.5
Kr—Xe	199	Rb ⁺ —Cs ⁺	104
Xe—Xe	293	Cs ⁺ —Cs ⁺	155

† The experimental values have been obtained, for the rare gases, from C. and M. Cuthbertson (1910), and for the ions in crystals, from K. Spangenberg (1923). For Li⁺, the theoretical value $\alpha = 0.0313 \times 10^{-24}$, calculated by Hassé (1930), has been used.

2—DIPOLE-QUADRIPOLE ENERGY

The most important term in the van der Waals energy of two atoms has been shown to be that which varies as R^{-6} , and when R is large it is unnecessary to consider other terms in the interaction energy. But for moderate values of R , at which the exchange forces between the atoms are still small, it is desirable to consider the second term proportional to R^{-8} , arising from the R^{-4} term in v_{pr} (equation (2)), and which may be regarded as the result of the interaction of oscillating dipoles in one atom with quadripoles in the other. The procedure by which we evaluate the corresponding term in the van der Waals energy is exactly similar to that used in finding the dipole-dipole term, though in practice it is much more laborious. We shall give a brief account of the analysis when only one parameter is used in the variation method; then the result for more parameters will be deduced without detailed calculation by analogy with the similar formulae for the dipole-dipole energy.

The energy is given by (3), when for v is substituted the dipole-quadrupole interaction, and is first of all reduced to a sum of simple integrals involving the electron wave functions, by integrating over the space of all electrons whose co-ordinates do not occur explicitly in the integrand and as usual, each term is averaged for all possible orientations of the two atoms relative to the line joining the nuclei. With the same notation as before we obtain

$$\left. \begin{aligned} \frac{1}{N_1! N_2!} \int \Psi^* v^2 \Psi d\tau &= \frac{1}{R^8} \sum_{\rho_1} \sum_{\rho_2} \{ (R^2)_{\rho_1} (R^4)_{\rho_2} + (R^4)_{\rho_1} (R^2)_{\rho_2} \}, \\ \frac{1}{N_1! N_2!} \sum_p \int \Psi^* (\text{grad}_p v)^2 \Psi d\tau &= \frac{1}{R^8} \sum_{\rho_1} \sum_{\rho_2} \{ 10(r^2)_{\rho_1 \rho_1} (R^2)_{\rho_2} + 3(R^4)_{\rho_2} \}, \\ \frac{1}{N_1! N_2!} \sum_r \int \Psi^* (\text{grad}_r v)^2 \Psi d\tau &= \frac{1}{R^8} \sum_{\rho_1} \sum_{\rho_2} \{ 10(R^2)_{\rho_1} (r^2)_{\rho_2 \rho_2} + 3(R^4)_{\rho_2} \}, \end{aligned} \right\} \quad (12)$$

where

$$(R^4)_{\rho_1} = (r^4)_{\rho_1 \rho_1} - 5 \sum_{\sigma_1 \neq \rho_1} \{ (xy)_{\sigma_1 \rho_1} (xy)_{\rho_1 \sigma_1} + (yz)_{\sigma_1 \rho_1} (yz)_{\rho_1 \sigma_1} + (zx)_{\sigma_1 \rho_1} (zx)_{\rho_1 \sigma_1} \}$$

and $(R^4)_{\rho_2}$ is a similar expression for the state ρ_2 in the second atom; $(R^2)_{\rho_1}$ and $(R^2)_{\rho_2}$ have been defined in (4).

After combining these results

$$\left. \begin{aligned} \bar{h} &= \frac{1}{R^8} \sum_{\rho_1} \sum_{\rho_2} \{ 2\lambda A_{\rho_1 \rho_2} + \frac{1}{2} \lambda^3 B_{\rho_1 \rho_2} \} \\ \text{in which} \quad A_{\rho_1 \rho_2} &= (R^2)_{\rho_1} (R^4)_{\rho_2} + (R^4)_{\rho_1} (R^2)_{\rho_2}, \\ B_{\rho_1 \rho_2} &= 10\{ (r^2)_{\rho_1 \rho_1} (R^2)_{\rho_2} + (R^2)_{\rho_1} (r^2)_{\rho_2 \rho_2} \} + 3\{ (R^4)_{\rho_1} + (R^4)_{\rho_2} \}, \end{aligned} \right\} \quad (13)$$

and when λ is chosen to make the energy a minimum,

$$\bar{h}_{\min} = -\frac{2}{R^6} \left\{ \sum_{\rho_1} \sum_{\rho_2} A_{\rho_1 \rho_2} \right\}^2 / \sum_{\rho_1} \sum_{\rho_2} B_{\rho_1 \rho_2}. \quad (14)$$

The result is expressed in this form so that the extension to the case when several parameters are used may be made easily. Thus when a parameter $\lambda_{\rho_1 \rho_2}$ is associated with the pair of states ρ_1, ρ_2 , the equation corresponding to (13) is

$$\bar{h} = \frac{1}{R^6} \sum_{\rho_1} \sum_{\rho_2} \{ 2\lambda_{\rho_1 \rho_2} A_{\rho_1 \rho_2} + \frac{1}{2} \lambda_{\rho_1 \rho_2}^2 B_{\rho_1 \rho_2} \} \quad (15)$$

(compare equations (5) and (9)), and hence, instead of (14),

$$\bar{h}_{\min} = -\frac{2}{R^6} \sum_{\rho_1} \sum_{\rho_2} A_{\rho_1 \rho_2}^2 / B_{\rho_1 \rho_2} \quad (16)$$

The dipole-quadrupole constant d_{12} is defined by $\bar{h}_{\min} = -d_{12}/R^6$, and is therefore

$$d_{12} = 2 \sum_{\rho_1} \sum_{\rho_2} A_{\rho_1 \rho_2}^2 / B_{\rho_1 \rho_2}. \quad (17)$$

This is the form of d_{12} which corresponds to (10) for the dipole-dipole constant c_{12} .

We shall not attempt to apply (17) to the general problem when the two interacting atoms are unlike, and in the following discussion the atoms are assumed to be alike. Moreover, it is assumed that the only important contribution to d_{12} arises from the interaction of an outer shell of ν_0 electrons (each with the same n, l quantum numbers) in one atom with the similar shell in the other atom. Then, since the terms $A_{\rho_1 \rho_2}^2 / B_{\rho_1 \rho_2}$ in equation (17) can be shown to depend only on the n, l quantum numbers of the states ρ_1 and ρ_2 the summation for different pairs of states can be made at once, with the result

$$d_{00} = 2\nu_0^2 A_{00}^2 / B_{00}, \quad (18)$$

where the suffix 0 refers to the states of the outer electrons, and

$$\left. \begin{aligned} A_{00} &= 2(R^2)_0 (R^4)_0, \\ B_{00} &= 20(r^2)_{00} (R^2)_0 + 6(R^4)_0. \end{aligned} \right\} \quad (19)$$

To the same approximation the dipole-dipole constant may be written, from equation (10),

$$c_{00} = \frac{2}{3} \nu_0^2 (R^2)_0^2, \quad (20)$$

and so combining equations (18), (19) and (20) a relation is obtained between c_{00} and d_{00}

$$d_{00} = 18c_{00} \frac{(R^4)_0^2}{10(r^2)_{00} (R^2)_0^2 + 3(R^2)_0 (R^4)_0}. \quad (21)$$

If c and d are now used to represent the van der Waals constants for two like atoms, including the contributions of inner electrons, then no appreciable error will be introduced by assuming that d and c are in the same ratio as d_{00} and c_{00} .

In order to get an idea of the relative magnitudes of the dipole-dipole and dipole-quadrupole energies for particular atoms, it is convenient to use the Slater wave functions of the form $r^{n^*-1} \exp(-n^*r/Z - S)$, where n^* is an effective quantum number and S a screening constant. We shall apply these functions to study two cases of immediate interest, in which the outer group contains either a pair of s electrons, as in He and normal Be, or two s electrons and six p electrons, as in the larger atoms and ions of rare-gas structure.

a—Outer group of two s electrons—When s states alone are occupied in the atoms, all exchange integrals vanish, and so

$$(R^2)_0 = (r^2)_{00}, \quad (R^4)_0 = (r^4)_{00}.$$

Also from the analytical form of the Slater functions, it is easily proved that

$$(r^4)_{00} = \frac{(n^* + \frac{3}{2})(n^* + 2)}{(n^* + \frac{1}{2})(n^* + 1)} (r^2)_{00}^2,$$

in which n^* depends only on the total quantum number n , hence from equation (21) it follows that

$$d = 18pc(r^2)_{00},$$

$$\text{where } p = \frac{x^2}{10 + 3x}, \quad x = \frac{(n^* + \frac{3}{2})(n^* + 2)}{(n^* + \frac{1}{2})(n^* + 1)}.$$

The usefulness of this relation is increased when $(r^2)_{00}$ is replaced by a quantity which can be derived from observable atomic properties, i.e. by χ_0 , the contribution of the outer group of electrons to the diamagnetic susceptibility. Thus when c and d are expressed in c.g.s. units

$$d = 18a_0^2 pc(r^2)_{00}, \quad (22)$$

and in the same units

$$\chi_0 = -\frac{Le^2 a_0^2}{6mc^2} \cdot \nu_0 (r^2)_{00}, \quad (23)$$

therefore, when $\nu_0 = 2$,

$$\begin{aligned} d &= -54 \frac{mc^2}{Le^2} \cdot pc \chi_0 \\ &= -3 \cdot 18 \times 10^{-18} (\chi_0 \times 10^6) pc. \end{aligned} \quad (24)$$

This result has been applied to the atoms of He, Li⁺ and Be, and the values of d are given in Table III below; in Table II are shown the values of p for total quantum numbers 1 to 5.

*b—Outer group of two s and six p electrons—*This case differs from the previous one in the occurrence of exchange integrals, the most important of which arise from the interaction of the *s* and *p* subgroups. For this reason, the presence of the *s* electrons cannot be disregarded, although their direct contribution to the van der Waals constants is negligible. The *p* electrons contribute nearly the whole of *c* and *d*, and the appropriate value of ν_0 is therefore six. There is no difficulty in reducing the exchange integrals to integrals involving the radial wave functions only; if the integration over the angular co-ordinates is completed and the exchange terms summed, it is found that†

$$(R^2)_0 = (r^2)_{00} - \frac{1}{2}(r)_{00}^2,$$

$$(R^4)_0 = (r^4)_{00} - \frac{3}{2}(r^2)_{00}^2,$$

provided that any *d* electrons present are neglected. Now from the form of the Slater functions, $(r^4)_{00} = x(r^2)_{00}^2$ as before, and

$$(r)_{00}^2 = y(r^2)_{00}, \text{ where } y = \frac{n^* + \frac{1}{2}}{n^* + 1};$$

consequently the relation (21) between the van der Waals constants becomes

$$d = 18qc(r^2)_{00},$$

where $q = (x - \frac{2}{3})^2/10(1 - \frac{1}{3}y)^2 + 3(1 - \frac{1}{3}y)(x - \frac{2}{3})$.

As before, χ_0 is substituted for $(r^2)_{00}$ from (23) with $\nu_0 = 6$; hence

$$\begin{aligned} d &= -18 \frac{mc^2}{Le^2} \cdot qc\chi_0 \\ &= -1.06 \times 10^{-18} (\chi_0 \times 10^4) qc. \end{aligned} \quad (25)$$

Values of *q* are given in Table II.

TABLE II—VALUES OF *p*, *q* FROM SLATER FUNCTIONS‡

<i>n</i>	1	2	3	4	5
<i>n</i> *	1	2	3	3.7	4.0
<i>p</i>	0.3571 (He)	0.2235 (Be)	0.1742 (Mg)	0.1556 (Ca)	0.1494 (Sr)
<i>q</i>	—	0.2563 (Ne)	0.1924 (Ar)	0.1675 (Kr)	0.1593 (Xe)

‡ The values of *n** are as given by Slater (1930).

It is possible to estimate the values of χ_0 from the observed susceptibilities and in fact this has already been done in Paper I for a number of atoms and

† The result for $(R^2)_0$ is simplified by the fact that Slater adopts the same radial wave function for the *ns* and *np* electrons.

ions. Their dipole-quadrupole constants can therefore be calculated from (24) and (25), using the dipole-dipole constants in Table I; the results are shown in Table III. The values for the negative ions F^- and Cl^- have been included, although the experimental data are unreliable. A few results by other writers are given for comparison; for example, Margenau (1931) has obtained for the dipole-quadrupole constant of two helium atoms a theoretical value 3.24×10^{-76} , which agrees well with 3.53×10^{-76} derived here from experimental data. Mayer (1933) has also estimated the van der Waals constants for the alkali and halogen ions in cubic crystals, basing his calculations on the observed ultra-violet absorption of the crystals of NaCl, KCl and KI. His values, the c and d constants, are rather lower than those obtained from atomic polarizabilities and susceptibilities, particularly for the smaller atoms, but they indicate that the dipole-quadrupole terms contribute to the crystal energy about 20 % of the dipole-dipole energy.

TABLE III—DIPOLE-QUADRIPOLE CONSTANTS

	$-\chi_0 \times 10^4$	$c \times 10^{60}$ erg. cm. ³	$d \times 10^{76}$ erg. cm. ⁵	$R_0 \times 10^8$ cm.	$d \times 10^{76}$ (Mayer)
He	1.90	1.63	3.53	2.94	(3.24)†
Li ⁺	0.70	0.0975	0.0775	1.8	0.03
Be	13.3‡	(498)	(4710)	6.15	—
F ⁻	7.6	25.0	52	2.85	17-23
Ne	5.37	7.48	10.9	2.4	—
Na ⁺	4.75	3.70	4.8	2.3	0.8
Cl ⁻	18.9	159	615	3.93	223-260
A	14.7	63.5	190	3.45	—
K ⁺	10.7	33.3	73	2.95	24.0
Kr	17.3	136	420	3.5	—
Rb ⁺	13.2	70.5	165	3.1	82
Cs ⁺	19.3	155	505	3.6	278

† The value of χ_0 for Be is calculated from the Fock field (D. R. and W. Hartree 1935).

‡ Margenau (1931).

In Table III are given the values of a quantity R_0 which probably have more significance than the dipole-quadrupole constants themselves. We define R_0 as the internuclear distance at which the dipole-quadrupole term contributes to the total van der Waals energy 25 % of the dipole-dipole term, i.e. $R_0^3 = 4d/c$, which by equations (24) and (25) is related very simply to the atomic structure. At distances less than R_0 the dipole-quadrupole term of course adds more than 25 %, but this percentage, though an arbitrary choice, is found to be a convenient one. From Table III it is seen that for a given nuclear separation the $d-q$ term increases in importance for the

heavier atoms (though helium is an exception), and is also more important for negative than for positive ions. In applications of the van der Waals constants, however, it is often more useful to know what correction should be applied to the dipole-dipole energy than to know the precise value of the dipole-quadrupole energy. For example, when the van der Waals energies are deduced from observed collision cross-sections, whether the results of molecular ray experiments or measurements of virial coefficients, the total energy is expressed in the form C/R^6 , and of this a proportion properly arises from the dipole-quadrupole and higher terms. It is therefore desirable to know what allowance should be made before comparing the observed values of c with theoretical values or those calculated from polarizabilities. Our values for R_0 now prove to be of some assistance in this. Thus at 10°C ., the effective collision diameter of argon atoms is about 10.7 \AA , and for neon atoms about 6.1 \AA , and on comparing these diameters, with R_0 it appears that in collisions the dipole-quadrupole term contributes very little to the total van der Waals energy, about $2\frac{1}{2}\%$ in argon and 4% in neon. But for helium atoms, from the effective collision diameter 2.97 \AA calculated by Massey and Mohr (1934), the dipole-quadrupole energy adds 25% to the dipole-dipole energy. At high temperature, the dipole-quadrupole terms will be relatively more important and at very low temperature, less so.

The calculation of the energy and other properties of crystals also requires a knowledge of van der Waals constants. The main contribution to the energy of crystals, such as the alkali halides, arises from the interaction of unlike ions, and in NaF, KCl, RbBr, and CsI crystals, the shortest distances between unlike ions are 2.31 , 3.14 , 3.44 and 3.95 \AA respectively. On comparing these separations with the values of R_0 in Table III, it appears that the dipole-quadrupole energy is 25 – 30% of the dipole-dipole energy. The negative ions are more widely spaced, but an allowance of 20% is still necessary, thus the F^- ions in NaF are 3.27 \AA apart and the value of R_0 is 2.85 \AA . The corresponding distances for the Cl^- ions in KCl are 4.44 and 3.95 \AA .

We shall conclude with a short discussion of the dipole-dipole constants as derived from three distinct sets of experimental data: polarizabilities, the dispersion formulae, and effective collision diameters. The results for the rare gases are shown together in Table IV; the polarizability values of $c \times 10^{40}$ are those calculated earlier in this paper; the dispersion values have been derived by London (1930, 1937) to whom is also due the theory relating the van der Waals energy to the dispersion; the collision values† have been obtained from the cross-sections for collisions between rare-gas and alkali

† A preliminary account has been given by H. S. W. Massey and the author (1936).

atoms observed by Rosin and Rabi (1935). For the alkali ions and negative ions F^- , Cl^- also, the dispersion results of Mayer (1933) are shown side by side with the polarizability results.

TABLE IV—DIPOLE-DIPOLE CONSTANTS, $c \times 10^{60}$ erg. cm.⁶

	Polariza- bility	Dis- persion	Collision cross- section		Polariza- bility	Dis- persion
He	1.63	1.23	1.32	Na ⁺	3.70	1.68
Ne	7.48	4.69	6.40	K ⁺	33.3	24.3
A	63.5	55.5	107	Rb ⁺	70.5	59.4
Kr	136	110	—	Cs ⁺	155	152
Xe	293	234	—	F ⁻	25.0	16.5-19.1
				Cl ⁻	159	116-130

It has already been emphasized that the relation between c and the polarizability leads to a value of c which should be regarded as an upper limit to the true dipole-dipole constant, and so it is not surprising that the polarizability values are always larger than the dispersion values. It should be mentioned that the dispersion values were derived on the assumption that the refractive index can be fairly well represented by a dispersion formula of one term only, i.e. that the absorption and emission is monochromatic, with the exception of the negative ions, for which Mayer has taken into account the continuous absorption spectrum. The effect of this more precise analysis was to double the constants for the negative ions; it is clear, however, that for heavy atoms, such as Kr and Xe, the assumption of monochromatic absorption is more or less valid. For Na⁺, K⁺ most of the discrepancy is probably due to Mayer having adopted rather smaller values for the polarizabilities of these ions than those used here,† and so the assumption is probably justified for these ions also. As mentioned earlier, the use of more accurate perturbed wave functions is likely to lower the polarizability values of c for light atoms; this is proved for helium by the theoretical value 1.324×10^{-60} which Hassé (1931) obtained with a more accurate variation method, and probably the neon constant could be reduced in the same way. On the other hand the dispersion value for neon might be increased by more exact analysis of the absorption.

The values derived from collision cross-sections, which have been corrected as suggested above for the dipole-quadrupole terms, are in excellent agreement for He and Ne, though the argon value is apparently high. The experi-

† The simple dispersion formula for c for like atoms is $c = \frac{1}{2} h \nu_0 \alpha^2$, where ν_0 is the characteristic frequency.

ments involved are difficult, but the results may be accepted as confirming those of other methods.

In conclusion, I wish to thank Professor J. E. Lennard-Jones, F.R.S., with whom these investigations were begun at Cambridge and Dr H. S. W. Massey for much helpful discussion.

SUMMARY

The method of a previous paper is extended to the polarization of an atom by another distant atom, and their mutual energy (the van der Waals energy) derived. The dipole-dipole constant is related to the atomic polarizabilities of the atoms, and the dipole-quadrupole constant also expressed in a simple way in terms of observable quantities. Dipole-dipole constants are calculated for rare-gas atoms and for alkali ions in crystals, and the results compared with those obtained in other ways. The importance of the dipole-quadrupole terms is discussed.

REFERENCES

- Cuthbertson, C. and M. 1910 *Proc. Roy. Soc. A*, **84**, 13.
Hartree, D. R. and W. 1935 *Proc. Roy. Soc. A*, **150**, 9.
Hassé, H. R. 1931 *Proc. Camb. Phil. Soc.* **27**, 66.
Kirkwood, J. G. 1932 *Phys. Z.* **33**, 57.
London, F. 1930 *Z. phys. Chem. B*, **11**, 222.
— 1937 *Trans. Faraday Soc.* **3**, 8.
Margenau, H. 1931 *Phys. Rev.* **38**, 747.
Massey, H. S. W. and Buckingham, R. A. 1936 *Nature, Lond.*, **138**, 77.
Massey, H. S. W. and Mohr, C. B. O. 1933 *Proc. Roy. Soc. A*, **141**, 434.
— — 1934 *Proc. Roy. Soc. A*, **144**, 188.
Mayer, J. E. 1933 *J. Chem. Phys.* **1**, 270.
Rosen, S. and Rabi, I. I. 1935 *Phys. Rev.* **48**, 373.
Slater, J. C. 1930 *Phys. Rev.* **36**, 57.
Spangenberg, K. 1923 *Z. Kristallogr.* **57**, 517.
-

Equilibrium Curve and Entropy Difference between the Supraconductive and the Normal State in Pb, Hg, Sn, Ta, and Nb

BY J. G. DAUNT AND K. MENDELSSOHN

Clarendon Laboratory, Oxford

(Communicated by F. A. Lindemann, F.R.S.—Received 29 January 1937)

The thermodynamical treatment of supraconductivity (Gorter and Casimir 1934*a, b*) is based on the fact that at the transition from the normal into the supraconductive state the induction becomes zero and it was assumed that the magnetic threshold curve, i.e. that curve in the (H, T) diagram that indicates the disappearance of the electric resistance, is also the equilibrium curve between the supraconductive and the normal state. It has to be remembered, however, that zero resistance and zero induction in the supraconductive state are not causally connected by electrodynamic theory, but that their coincidence must as yet be considered as a purely empirical fact. It is therefore not possible to deduce the existence of zero induction from the observation of zero resistance and vice versa. It is, for example, known that zero induction and zero resistance do not coincide in supraconductive alloys (Mendelssohn 1935) and preliminary experiments on tantalum (Mendelssohn and Moore 1936) have shown that even in pure metals the threshold values obtained in conductivity experiments cannot be used for an evaluation of thermodynamical quantities.

In two previous communications (Keeley and Mendelssohn 1936; Mendelssohn 1936) two methods (later called 1 and 2) for the determination of the induction in supraconductors were described. With these methods the equilibrium curves of extremely pure Pb, Hg, Sn, Ta, and Nb were determined in the temperature interval between 1.5° and 4.2° K.

As the entropy difference between the supraconductive and the normal state seems to be the most interesting quantity that can be calculated from our results, curves for the variation of this quantity with the temperature have also been included in this paper. Furthermore, this entropy difference gives a lower limit for the value of the electronic entropy in the normal state, which has so far only been determined in a few metals by a direct measurement of the specific heats. The results obtained in the latter way are, however, necessarily obscured to some extent by the influence of the vibrations of the crystal lattice, which cannot always be calculated

very accurately, whereas such a correction does not enter into determinations from the equilibrium curves, as described in this paper.

It would take too long in this paper to attempt to compare the equilibrium values with the threshold values which we have found in conductivity experiments or to examine their relation to the threshold values determined at Leiden and elsewhere. Any deviation between threshold curve and the equilibrium curve does not affect the thermodynamical quantities as these are solely based on the latter. Discrepancies between the values obtained in induction and resistance measurements therefore do not affect our conclusions and their examination may be relegated to a later publication.

Great care was taken to employ only the purest substances, as this was found to be necessary in order to ensure a reversible change to zero induction at the equilibrium curve. A short discussion of the reversibility observed is given in each particular case.

SUBSTANCES

Tin—The specimen employed was made out of "H.S." tin from Adam Hilger Ltd. (Laboratory number 10,000). The purity of the material is given by the manufacturers as 99.996 %. The induction in spheres cast *in vacuo* was determined and the equilibrium curve was taken from the values at which the induction became normal and by multiplying the field at which first flux penetrated into the specimen by $3/2$. The reversibility had been tested by previous experiments (Mendelssohn, Daunt and Pontius 1936) and found to be practically ideal. The amount of flux "frozen in" at zero field in sphere was not greater than 1 %.

Lead—A lead sphere was cast *in vacuo* out of "H.S." lead from Adam Hilger Ltd (Laboratory number 8334). The purity was given as higher than 99.999 %. The reversibility of this material had been tested by Shoenberg (1936) and found to be very good. This result was confirmed by experiments carried out on our sample with both methods. In determining the equilibrium curve, however, we observed some hysteresis effects in so far as frequently the field strength at which supraconductivity was destroyed did not coincide with the field strength at which the first decrease in induction was observed, when the external field was lowered.* This

* In a different set of experiments, in which the shielding of the persistent currents in tin and lead near the equilibrium curve was investigated by a special method, it was found that, under certain conditions, the shielding still existed at values of field and temperature, which lay considerably above the equilibrium curves given in this paper and the threshold values determined in Leiden. A detailed report on these phenomena will be given in a later paper.

behaviour was especially marked at lower temperatures. As in previous experiments the final value of induction was only attained after some time (Mendelssohn 1936, p. 563). It was found, however, that, if one waited long enough, the supraconductive state disappeared and re-emerged at the same field strength. Only values obtained in this way have been included in our results.

Mercury—Purest mercury from Hopkin and Williams Ltd. was used, which was once more distilled before being brought into a spherical glass container. The reversibility had been found to be very good in previous experiments (Mendelssohn 1936). The flux "frozen in" amounted in the sphere employed in this investigation to only 1.3 %.

Tantalum—The specimen employed was a rod of "H.S." tantalum kindly lent to us by Adam Hilger Ltd. (Laboratory number 10,879). The sample had a purity of over 99.95 %. The equilibrium curve was determined by both methods which gave identical results. The reversibility was found to be very good near the transition point, but decreased at lower temperatures. This fact has probably to be ascribed to the existence of regions of higher threshold value in the specimen which form a sort of supraconductive sponge (Mendelssohn 1935) and prevent some of the lines of force from escaping. Such an effect would be more noticeable at lower temperatures where the threshold value is higher, and the discrepancy between the threshold values of the different parts of the specimen would accordingly be greater. The same behaviour has been observed in supraconducting alloys and previous experiments on tantalum (Mendelssohn and Moore 1936) have shown that the tendency to form a supraconductive sponge is very strong in this metal. The rod we used is actually the first specimen in which reversibility has been observed at all. The curve obtained therefore cannot be considered to give equilibrium values with the same accuracy as those of tin, lead and mercury, but our experiments seem to indicate that the actual equilibrium curve will not show values very different from those given in this paper.

Niobium—A rod of "H.S." niobium of about 99.8 % purity was kindly lent to us by Adam Hilger Ltd. (Laboratory number 8290). The equilibrium values of this substance, which was investigated by method 2, were found to be very high. The fields necessary had to be produced by a current taken from a generator which was, however, not constant enough to enable very accurate results to be obtained. The values given can therefore only be considered as preliminary. The change of induction was found to be partly reversible, and the flux "frozen in" at zero field amounted to about 20 %.

RESULTS

Results are given in fig. 1 and Table I, the latter representing values from the smoothed curve. The error arising from the determination of field and temperature of an individual point on the equilibrium curve was smaller than 1 % (see also Mendelssohn 1936). An additional uncertainty is introduced by hysteresis phenomena (Mendelssohn and Pontius 1936*a*) which is expressed by a scattering of the points. We were able, however, to reduce this error by long waiting at each point, making the specimen of suitable geometrical shape (Mendelssohn and Pontius 1936*b*) and by taking a great number of readings so that the total error is never greater than 1 %, except for niobium, where the error may be 2 %.

TABLE I—EQUILIBRIUM FIELDS (GAUSS)

Element	4.2°	4.0°	3.5°	3.0°	2.5°	2.0°	1.5°
Tin	—	—	30.5	95.5	152	206	246
Mercury	0	34	121	200	269	320	359
Lead	552	573.5	628	675	717	749	772
Tantalum	74	139	298	454	588	716	825
Niobium	2040	2085	2180	2270	2360	2440	2510

DISCUSSION

All these equilibrium curves are somewhat similar though not exactly identical in shape, and it does not seem possible to represent them by a simple function. None of these curves is a parabola, which means that Kok's (Kok 1934) relation between the supraconductive and the normal entropies does not hold for these metals and that his considerations cannot be applied except perhaps as a first approximation. It is interesting to note that the equilibrium curves of the "soft" superconductors (i.e. those of low Debye Θ , which stand in the B groups of the periodic system) are all about equally steep and do not intersect, but that they are intersected by the equilibrium curves of tantalum, which like niobium is a "hard" superconductor.

The entropy difference ΔS can be derived from the equilibrium curve by the following equation

$$\Delta S = \frac{A}{4\pi d} \times H \times \frac{dH}{dT},$$

H is the field at which the two states are in equilibrium and A/d is the atomic volume of the substance. The values of ΔS obtained in this way are given in fig. 2 and Table II. In all cases it can be seen that with decreasing

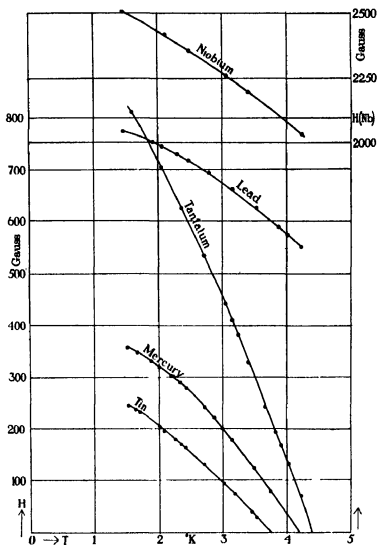


FIG. 1—Equilibrium curves.

TABLE II— ΔS (cal. $\times 10^4$)

Element	4.0°	3.75°	3.5°	3.25°	3.0°	2.75°	2.5°	2.25°	2.0°	1.75°
Tin	—	—	1.25	2.51	3.44	4.34	5.09	5.69	6.3	6.1
Mercury	1.51	2.75	5.18	6.75	7.92	8.79	8.84	8.64	7.95	7.15
Lead	21.4	21.5	21.4	21.0	20.7	19.6	18.3	15.5	14.0	11.1
Tantalum	8.8	—	18.5	—	27.6	—	34.7	35.8	35.5	33.6
Niobium	86.8	—	—	—	88.0	—	—	—	80.3	—

temperature the entropy difference increases continuously from the value zero at the transition point and passes through a maximum at about half the transition temperature. An extrapolation of ΔS to absolute zero is possible, if we assume that the entropy of the supraconductive state is already very small and that of the normal is a linear function of T according to Sommerfeld's formula for the electronic specific heats. This latter conclusion is reiterated by Sommerfeld in a very recent paper (Sommerfeld 1937) in which the linear increase of the electronic specific heat with

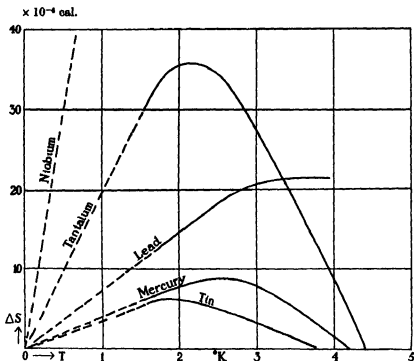


FIG. 2.—Difference of entropy.

temperature is asserted to be generally valid. The results, in the cases of mercury and lead particularly, indicate that a linear decrease of ΔS at low temperatures really occurs. The extrapolation of ΔS to absolute zero for tin is far less reliable as at the lowest temperature to which we can evaluate our results the maximum in ΔS has only just been passed. The ΔS curves for tantalum and niobium can also be traced well below the maximum; they are, however, owing to the above-mentioned experimental difficulties, less accurate than the curves for the "soft" superconductors.

It is now possible to draw certain conclusions about the electronic entropies in the normal state. In this state the total electronic entropy

can be expressed numerically as the sum of the entropy of the supraconductive electrons plus the excess entropy of the normal over the supraconductive state. The first term may be small but of course cannot be negative, so that ΔS gives a lower limit for the total electronic entropy in the normal state. In Table III the entropy differences derived from our

TABLE III— $T = 1.5^\circ \text{ K.}$

Element	n	$S_{\text{theor.}}$	$\Delta S_{\text{exp.}}$
Tin	4	$4.72 \times 10^{-4} \text{ cal.}$	$5.2 \times 10^{-4} \text{ cal.}$
Mercury	2	$3.53 \times 10^{-4} \text{ cal.}$	$5.62 \times 10^{-4} \text{ cal.}$
Lead	4	$5.31 \times 10^{-4} \text{ cal.}$	$10.6 \times 10^{-4} \text{ cal.}$

values at 1.5° K. are given together with the total electronic entropies of the normal metal calculated from Sommerfeld's formula

$$c = 3.26 \times 10^{-5} \times n^{\frac{1}{2}} \times \left(\frac{A}{d}\right)^{\frac{1}{2}} \times T$$

(except for tantalum and niobium) on the assumption that the number of electrons is equal to the valency.

It is interesting to note that in all three cases the entropy difference is found to be higher than the total entropy calculated from Sommerfeld's formula. Whereas the discrepancy is small (and, as mentioned above, not very accurately defined) in the case of tin, it is strongly marked in mercury and lead. In the latter case our value is twice the theoretical one, in other words, even if we assume the entropy of the supraconductive state to be already zero at 1.5° K. the entropy of the normal must be about twice as great as the value derived from Sommerfeld's formula.

A further interesting feature of our results is the very high values of ΔS for tantalum and niobium, which suggest, as in the case of lead and mercury, very high electronic entropies for the normal state. It has been pointed out by Mott (1936) that entropies of this order may be expected in the so-called transition metals and direct evidence of this has been found in the specific heats of nickel (Keesom and Clark 1935), platinum (Kok and Keesom 1936) and palladium (Pickard 1936). In these metals the contribution of the "free" (s -) electrons to the specific heat is small compared with the contribution by the electrons in a partly unfilled lower (d -) shell. Tantalum and niobium, in view of their position in the periodic system, can also be counted in the category of the transition metals.*

The decrease in entropy which occurs when a metal is brought from the normal into the supraconductive state indicates that the electrons pass

* This was kindly pointed out to us by Professor Mott.

into an arrangement of higher order than that corresponding to the so-called "Fermi gas".* As the contribution of the electrons in the (*d*-) shell to the electronic entropy in tantalum and niobium vanishes at the transition to the superconductive state it is evident that they must take part in the rearrangement. This means that when these metals become superconductive not only the free electrons but also those in the partly unfilled lower shell are affected.†

Another thermodynamic quantity that can be calculated from the equilibrium curves is the difference between the specific heat in the superconductive and in the normal state. In fig. 3 this difference (Δc) for mercury and tantalum is plotted. On the assumption that the specific heat of the normal state changes linearly with temperature an estimate for the specific heat of the superconductive electrons can be derived from the difference between Δc and the tangent to the Δc curve at absolute zero. This tangent could be drawn with some accuracy for mercury though the error is much greater for tantalum. It is clear, however, from the Δc curve that the resultant curve for the electronic specific heat in the superconductive state will not be a simple function. The reason for this may of course be the deviation of the normal specific heat from a straight line; again it may be due to changes which perhaps occur in the superconductive state at different temperatures. Finally, the possibility cannot be excluded that a slight change in the lattice entropy occurs at the transformation from the normal to the superconductive state although this seems rather improbable.

More information on these questions might be obtained by direct determination of the electronic specific heats. Our results indicate that the additional specific heat due to the electrons will be so high in tantalum and niobium that it should be possible to separate it from the specific heat of the lattice with some accuracy. It is also of importance to determine the threshold curves in the temperature region below 1° K. and to extend the investigations to a greater number of substances.

Finally it might be mentioned that tantalum and niobium seem to be very suitable substances for the production of very low temperatures by the method of adiabatic magnetisation described previously (Mendelssohn and Moore 1934; Mendelssohn, Daunt and Pontius 1936).

* The results of recent experiments on the absorption of infra-red light in superconductors (J. G. Daunt, T. C. Keeley and K. Mendelssohn 1937) might possibly be interpreted as direct evidence for such a state of higher order.

† In this connexion it must be remembered that there is no sharp distinction between the "free electrons" and those in the unfilled *d* band, the difference being merely in the "effective mass" of the two kinds.

Our thanks are due to Professor F. A. Lindemann, F.R.S., for the interest he has taken in the research, to Mr R. B. Pontius for help during the experiments, and to Imperial Chemical Industries, Ltd., who made research possible for one of us.

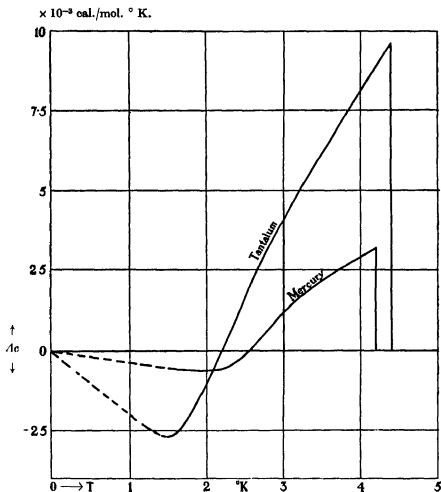


FIG. 3.—Difference in specific heats.

SUMMARY

The curves in the (H, T) diagram showing the equilibrium between the supraconductive and the normal state have been determined for Pb, Hg, Sn, Ta, and Nb from measurements of the magnetic induction.

The entropy differences between the two states have been calculated and the following conclusions were drawn:

(1) that the electronic entropies of non-supraconductive Pb and Hg are considerably higher than those given by Sommerfeld's formula on the assumption that the number of free electrons is equal to the valency,

(2) that in Ta and Nb not only the valency electrons but also the electrons in the partly unfilled lower shell take part in the formation of the supraconductive state.

The difference in specific heat between the two states has been calculated for Hg and Ta.

REFERENCES

- Daunt, J. G., Koeley, T. C. and Mendelssohn, K. 1937 *Phil. Mag.* **23**, 264-71.
 Gorter, C. J. and Casimir, H. 1934a *Physica, 's Grav.*, **1**, 305-20.
 — — 1934b *Z. techn. Phys.* **15**, 539-42.
 Koeley, T. C. and Mendelssohn, K. 1936 *Proc. Roy. Soc. A*, **154**, 378-85.
 Keesom, W. H. and Clark, S. W. 1935 *Physica, 's Grav.*, **2**, 513-20.
 Kok, J. A. 1934 *Physica, 's Grav.*, **1**, 1103-6.
 Kok, J. A. and Keesom, W. H. 1936 *Physica, 's Grav.*, **3**, 1035-45.
 Mendelssohn, K. 1935 *Proc. Roy. Soc. A*, **152**, 34-41.
 — — 1936 *Proc. Roy. Soc. A*, **155**, 558-70.
 Mendelssohn, K., Daunt, J. G. and Pontius, R. B. 1936 *VII Congrès Int. du Froid*, The Hague.
 Mendelssohn, K. and Moore, Judith R. 1934 *Nature, Lond.*, **133**, 413.
 — — 1936 *Phil. Mag.* **21**, 532-44.
 Mendelssohn, K. and Pontius, R. B. 1936a *Physica, 's Grav.*, **3**, 327-31.
 — — 1936b *Nature, Lond.*, **138**, 29-30.
 Mott, N. F. 1936 *Proc. Roy. Soc. A*, **152**, 42-6.
 Pickard, G. L. 1936 *Nature, Lond.*, **138**, 123.
 Shoenberg, D. 1936 *Proc. Roy. Soc. A*, **155**, 712-26.
 Sommerfeld, A. 1937 *Ann. Phys., Lpz.*, **28**, 1-10.
-

On Antiplane Stress in an Elastic Solid

By L. N. G. FILON, F.R.S.

(Received 12 January 1937)

1—It is usual to apply the term “plane stress” to a stress system in an elastic solid in which all the stress components other than those acting solely parallel to a fixed plane (which may be taken as that of x, y) vanish identically.

Thus the characteristic of plane stress is that

$$\hat{xz} = 0, \hat{yz} = 0, \hat{zz} = 0,$$

but $\hat{xx}, \hat{yy}, \hat{xy}$ do not in general vanish.

We may use the term “antiplane stress” to denote a stress system which is complementary to plane stress, that is, one in which

$$\hat{xx} = 0, \hat{yy} = 0, \hat{xy} = 0, \quad (1)$$

but $\hat{xz}, \hat{yz}, \hat{zz}$ do not in general vanish.

It is well known that the usual solutions for Euler-Bernoulli flexure, for de Saint-Venant flexure and torsion, as well as that for simple longitudinal tension, belong to the antiplane type.

There is, however, one class of solutions of this type which appears to have generally escaped notice, although it leads to examples of stress not without practical interest.

The object of the present paper is to develop a theory of this class of solutions.

2—It can be shown that if any set of stresses satisfy the equations

$$(1 + \eta) \nabla^2 \hat{p}q + \partial^2 \Theta / \partial p \partial q = 0, \quad (2)$$

where p, q are any of x, y, z ; η is Poisson's ratio; and $\Theta \equiv \hat{xx} + \hat{yy} + \hat{zz}$; and if, in addition, these stresses satisfy the equations

$$\partial \hat{px} / \partial x + \partial \hat{py} / \partial y + \partial \hat{pz} / \partial z = 0, \quad (3)$$

then the set of stresses are possible stresses in an elastic solid in equilibrium without body force.

If, in equations (3), we substitute the conditions (1), we obtain

$$\partial \hat{xz} / \partial z = 0, \partial \hat{yz} / \partial z = 0, \partial \hat{xz} / \partial x + \partial \hat{yz} / \partial y + \partial \hat{zz} / \partial z = 0. \quad (4)$$

Differentiating the last of (4) with respect to z and using the first two, we have

$$\partial^2 \widehat{z\bar{z}} / \partial z^2 = 0. \quad (5)$$

Again, substituting from (1) into (2), we have, since $\Theta = \widehat{z\bar{z}}$,

$$\partial^2 \widehat{z\bar{z}} / \partial x^2 = 0, \quad \partial^2 \widehat{z\bar{z}} / \partial y^2 = 0, \quad \partial^2 \widehat{z\bar{z}} / \partial x \partial y = 0, \quad (6)$$

and

$$(1 + \eta) \nabla_2^2 \widehat{x\bar{z}} + \partial^2 \widehat{z\bar{z}} / \partial x \partial z = 0, \quad (7.1)$$

$$(1 + \eta) \nabla_2^2 \widehat{y\bar{z}} + \partial^2 \widehat{z\bar{z}} / \partial y \partial z = 0, \quad (7.2)$$

$$\nabla_2^2 \widehat{z\bar{z}} = 0, \quad (7.3)$$

where $\nabla_2^2 \equiv \partial^2 / \partial x^2 + \partial^2 / \partial y^2$.

(7.3) gives nothing new, being already satisfied in virtue of (6). (5) and (6) lead to

$$\widehat{z\bar{z}} = (A_0 + A_1 z)x + (B_0 + B_1 z)y + Pz + Q. \quad (8)$$

Now the solutions which are derived from A_0 , B_0 and A_1 , B_1 reduce to the usual Euler-Bernoulli and de Saint-Venant solutions for flexure.

The solution corresponding to Q is the ordinary solution for a bar under tension.

There is likewise a solution corresponding to $\widehat{z\bar{z}} = 0$, which is the standard solution for torsion.

The solutions corresponding to

$$\widehat{z\bar{z}} = Pz \quad (9)$$

do not appear to have received consideration.

It is this class of solution which we now propose to examine. Equations (4), (7.1), (7.2) and (9) give, $\widehat{x\bar{z}}$ and $\widehat{y\bar{z}}$ being functions of x , y only,

$$\partial \widehat{x\bar{z}} / \partial x + \partial \widehat{y\bar{z}} / \partial y + P = 0, \quad \nabla_2^2 \widehat{x\bar{z}} = 0, \quad \nabla_2^2 \widehat{y\bar{z}} = 0,$$

$$\text{whence} \quad \widehat{x\bar{z}} = -\partial \psi / \partial y - \frac{1}{2}Px, \quad \widehat{y\bar{z}} = \partial \psi / \partial x - \frac{1}{2}Py, \quad (10)$$

and

$$(\partial / \partial x) \nabla_2^2 \psi = 0, \quad (\partial / \partial y) \nabla_2^2 \psi = 0,$$

i.e. $\nabla_2^2 \psi = C$, a constant.

We may write $\psi + \psi_1$ instead of ψ and choose

$$\nabla_2^2 \psi_1 = C, \quad \nabla_2^2 \psi = 0,$$

and consider separately $\widehat{x\bar{z}}$, $\widehat{y\bar{z}}$ given by (10), where ψ is now a harmonic function, and $\widehat{x\bar{z}}_1$, $\widehat{y\bar{z}}_1$ given by

$$\widehat{x\bar{z}}_1 = -\partial \psi_1 / \partial y, \quad \widehat{y\bar{z}}_1 = +\partial \psi_1 / \partial x,$$

together with $\widehat{z\bar{z}}_1 = 0$. The solution $\widehat{x\bar{z}}_1$, $\widehat{y\bar{z}}_1$, $\widehat{z\bar{z}}_1$ reduces to the torsion solution and may be omitted.

Let now ϕ be the harmonic function of which ψ is the conjugate. Then the type of solution which we shall consider is given by

$$\widehat{xz} = -\partial\phi/\partial x - \frac{1}{2}Px, \quad \widehat{yz} = -\partial\phi/\partial y - \frac{1}{2}Py, \quad \widehat{zz} = Pz. \quad (11)$$

This satisfies all the stress equations inside the body, provided continuity exists.

The integration of the differential equation for the lines of resultant shear in a cross-section is difficult except in special cases, but their orthogonal curves are obtained at once, for they satisfy the equation

$$\widehat{xz} dx + \widehat{yz} dy = 0,$$

or
$$d\phi + \frac{1}{2}P(x dx + y dy) = 0,$$

of which the integral is
$$\phi + \frac{1}{4}P(x^2 + y^2) = \text{const.},$$

and is known so soon as ϕ is known.

3—Let us now consider the conditions at the boundary of a cylinder of any shape, whose axis is parallel to the axis of z .

In the other problems of the antiplane type (tension, torsion and flexure) the curved surface of the cylinder is taken free from traction.

If we do this in the present case, a singularity must inevitably appear inside, for otherwise the normal tractions across the plane ends (say $z=0$ and $z=l$) would be unbalanced.

Let us nevertheless contemplate this possibility and take for the boundary condition the vanishing of the traction. Then

$$l\widehat{xz} + m\widehat{yz} = 0,$$

where (l, m) are the direction cosines of the outwards normal n to the boundary. This gives

$$-\partial\phi/\partial n = -\partial\psi/\partial s = \frac{1}{2}P(lx + my),$$

and since $l = \partial y/\partial s$, $m = -\partial x/\partial s$, s being the arc of the boundary,

$$lx + my = (x dy - y dx)/ds = r^2 \partial\theta/\partial s = 2\partial A/\partial s,$$

where A is the area swept out by the radius vector from the origin. Thus the boundary condition is

$$-\partial\phi/\partial n = -\partial\psi/\partial s = P\partial A/\partial s, \quad (12)$$

leading to

$$\psi + AP = \text{const.} \quad (13)$$

over the boundary.

It is then clear that ψ has a cyclic constant $-A_0P$ (where A_0 is the total area) after a complete circuit of the boundary, so that, in the two-dimen-

sional hydrodynamic analogue, where ϕ , ψ are the velocity-potential and stream function, there must be sources inside the boundary of total strength $A_0 P$. Subject to this overriding condition the position and strength of the individual sources may be arbitrarily prescribed.

It should be noted that the above result is independent of whether the cross-section is simply connected or not, as is obvious when we write the condition in the form

$$-\int (\partial\phi/\partial n) ds = PA_0, \quad (14)$$

the integral being taken over the complete boundary, with the usual sign conventions. The constant on the right-hand side of (13) may, however, be different over separate boundaries.

The meaning of the cyclic constant in the elastic problem is easily found. For, if we consider the stress $\hat{n}z$ across any closed contour s in the xy plane, we have, for the total force Z exerted by the material inside s upon the material outside,

$$\begin{aligned} Z &= -\int (l\hat{x}z + m\hat{y}z) ds \\ &= \int (d\psi + \frac{1}{2}Pr^2 d\theta) = Cy(\psi) + A_s P, \end{aligned}$$

where $Cy(\psi)$ denotes the cyclic constant of ψ after describing the contour and A_s is the area of the contour s . If now this contour is made to shrink into the point O , we have a concentrated force (per unit length of the cylinder)

$$Z = Cy(\psi, O) = K, \text{ say.}$$

We may write $\psi = K\theta/2\pi + \psi_1$, $\phi = (K \log r)/2\pi + \phi_1$, (15)

so that ϕ_1 , ψ_1 are conjugate functions which have no singularity at O . Thus a simple two-dimensional source in the hydrodynamic analogue corresponds to a uniform internal longitudinal pull of amount K per unit length, such as might be exerted by a "friction tube" through O .

Obviously, if there are a number of such singularities

$$\Sigma K + A_0 P = 0,$$

and we then write

$$\psi = \Sigma K\theta/2\pi + \psi_1, \quad \phi = \Sigma (K \log r)/2\pi + \phi_1.$$

4—The functions ϕ_1 , ψ_1 (and also ϕ , ψ) may be obtained, so soon as the Green's functions G , H for the given boundary are known.

Define G as usual from the conditions:

$G = 0$ over the boundary s of the cylinder;

$\nabla^2 G = 0$ inside the material;

$G \rightarrow \log r_P$ in the neighbourhood of an arbitrary point P , where r_P is the distance from P .

For simplicity we shall consider the case where there is only one singularity O inside, the argument being easily generalized. We have then

$$K = -A_0 P,$$

and, on the boundary, $\psi_1 = P(A_0\theta/2\pi - A)$,

ψ_1 being throughout a one-valued and continuous harmonic function.

We then obtain, using Green's theorem in the usual manner,

$$2\pi\psi_1 = \int \psi_1 (\partial G / \partial n) ds,$$

which determines the values of ψ_1 at P .

Substituting $\psi_1 = PA_0\theta/2\pi + \psi$,

$$2\pi\psi = \int \psi (\partial G / \partial n) ds + (PA_0/2\pi) \int \theta (\partial G / \partial n) ds - PA_0\theta.$$

But

$$\begin{aligned} \int \theta (\partial G / \partial n) ds &= \int \{ \theta (\partial G / \partial n) - G (\partial \theta / \partial n) \} ds \\ &= \int \{ \theta (\partial G / \partial \nu_0) - G (\partial \theta / \partial \nu_0) \} d\sigma_0 + \int \{ \theta (\partial G / \partial \nu) - G (\partial \theta / \partial \nu) \} d\sigma \\ &\quad + \int \{ \theta (\partial G / \partial \nu_\beta) - G (\partial \theta / \partial \nu_\beta) \} d\beta, \end{aligned}$$

where σ_0 , σ are small circles surrounding the points O and P and β is a barrier drawn from σ_0 to the point of s , where the discontinuity in A and θ is arbitrarily located, ν_0 , ν and ν_β are the normals to σ_0 , σ and β , both faces of the barrier being included in the integration.

Remembering that $\xi \log \xi \rightarrow 0$ when $\xi \rightarrow 0$, the right-hand side of the last-written equation becomes

$$2\pi\theta - 2\pi \int (\partial G / \partial \nu_\beta) d\beta,$$

where now $d\beta$ is a *simple* element of the barrier and $d\nu_\beta$ is an element of its normal in the sense of increasing θ .

Hence, finally,

$$\psi = (1/2\pi) \int \psi (\partial G / \partial n) ds - (PA_0/2\pi) \int (\partial G / \partial \nu_\beta) d\beta. \quad (16)$$

Similarly, defining H from the conditions

$\partial H / \partial n = 0$ over the boundary of the cylinder;

$\nabla^2 H = 0$ inside the cylinder;

$H \rightarrow \log r_C - \log r_P$, in the neighbourhood of points C and P , we obtain from Green's theorem, where

$$\phi_1 = \phi + (A_0 P / 2\pi) \log r,$$

$$\begin{aligned} \phi_1(P) &= \phi_1(C) + (1/2\pi) \int H(\partial \phi_1 / \partial n) ds \\ &= \phi_1(C) + (1/2\pi) \int H(\partial \phi / \partial n) ds + (PA_0/4\pi^2) \int H(\partial \log r / \partial n) ds. \end{aligned}$$

Evaluating the last-written integral as before, remembering that H has a singularity at C , as well as at P , and that $\log r$ is one-valued, so that no barrier is required, we find

$$\int H(\partial \log r / \partial n) ds = 2\pi H(O) - 2\pi \log OC + 2\pi \log OP,$$

whence, substituting into the previous equation,

$$\phi(P) = \phi(C) + (PA_0/2\pi) H(O) + (1/2\pi) \int H(\partial \phi / \partial n) ds.$$

The term $\phi(C)$, being independent of P , may be omitted without loss of generality, but $H(O)$ generally involves both C and P and so must be retained. We have thus

$$\phi = (PA_0/2\pi) H(O) + (1/2\pi) \int H(\partial \phi / \partial n) ds. \quad (17)$$

5.—To calculate the displacements in this strain, we have

$$2\mu(\partial u / \partial x) + \lambda \delta = 0, \quad 2\mu(\partial v / \partial y) + \lambda \delta = 0, \quad 2\mu(\partial w / \partial z) + \lambda \delta = Pz,$$

where $\delta = \partial u / \partial x + \partial v / \partial y + \partial w / \partial z$, so that $(3\lambda + 2\mu)\delta = Pz$, and the following particular integrals for u, v are obtained:

$$2\mu u = -\lambda Pxz / (3\lambda + 2\mu), \quad 2\mu v = -\lambda Pyz / (3\lambda + 2\mu). \quad (18)$$

These clearly satisfy the condition $\widehat{xy} = 0$.

Further $2\mu w = [(\lambda + \mu) Pz^2 / (3\lambda + 2\mu)] + 2W(x, y),$

whence $\widehat{xz} = \partial W / \partial x - \frac{1}{2} \lambda Px / (3\lambda + 2\mu) = -\partial \phi / \partial x - \frac{1}{2} Px,$

$\widehat{yz} = \partial W / \partial y - \frac{1}{2} \lambda Py / (3\lambda + 2\mu) = -\partial \phi / \partial y - \frac{1}{2} Py,$

leading to

$$\partial(W + \phi)/\partial x = -(\lambda + \mu) Px/(3\lambda + 2\mu), \quad \partial(W + \phi)/\partial y = -(\lambda + \mu) Py/(3\lambda + 2\mu),$$

or

$$W = -\phi - \frac{1}{2}(\lambda + \mu) P(x^2 + y^2)/(3\lambda + 2\mu),$$

and therefore

$$2\mu w = [(\lambda + \mu) P\{z^2 - (x^2 + y^2)\}/(3\lambda + 2\mu)] - 2\phi. \quad (19)$$

Since (18) and (19) lead to the required stresses, they give the values of the displacements, apart from an arbitrary rigid-body displacement.

We note from (19) that all cross-sections suffer identical transverse distortion, their relative longitudinal displacements as a whole being given by

$$2\mu w_0 = (\lambda + \mu) Pz^2/(3\lambda + 2\mu). \quad (20)$$

6—An important particular case arises when a singularity is on the boundary. Taking the origin O at the singularity and selecting O as the point of discontinuity for the multiple-valued functions A and θ , we have

$$Cy(A) = A_0,$$

but

$$Cy(\theta) = \pi$$

instead of 2π as before.

We find again, integrating round a semicircular notch surrounding O , that

$$Z = -PA_0,$$

giving a concentrated longitudinal shear Z applied to the boundary of the cylinder along the generator through O .

We then write here

$$\phi = -(PA_0/\pi) \log r + \phi_1, \quad \psi = -PA_0\theta/\pi + \psi_1,$$

and ϕ_1, ψ_1 are throughout continuous and one-valued up to and including the boundary and are determined in the usual manner.

In practice a stress system of this kind can be applied to the cylinder either by means of a series of small projecting studs subjected to appropriate transverse forces parallel to the cylinder axis, or by longitudinal friction.

7—The general method by Green's functions, given in § 4, is frequently either inapplicable (Green's functions being known only for a few contours) or cumbersome. Thus, for the important case of a solid circular cylinder, it is more convenient to proceed directly.

Consider first the circular cylinder of radius a (fig. 1) with a simple singularity O on the boundary. Let C be the centre of the circular cross-section. Take OC as axis of x , Q any point of the boundary, angle $QOC = \theta$ and the sector $QOA = A$, then

$$A = a^2\theta + \frac{1}{2}ay = A_0\theta/\pi + \frac{1}{2}ay.$$

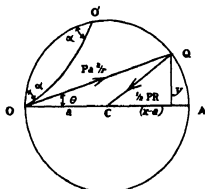


FIG. 1

The boundary condition for ψ_1 is

$$\psi_1 = P(A_0\theta/\pi - A) = -\frac{1}{2}Pay,$$

and, since the right-hand side is already a harmonic function, it gives the value of ψ_1 throughout.

Thus (r, θ) being polar co-ordinates referred to pole O and initial line OC

$$\psi = -Pa^2\theta - \frac{1}{2}Pay, \quad (21.1)$$

$$\text{and} \quad \hat{x}\hat{z} = Pa^2x/r^2 + \frac{1}{2}P(a-x), \quad (21.2)$$

$$\hat{y}\hat{z} = Pa^2y/r^2 - \frac{1}{2}Py. \quad (21.3)$$

The shears across the section are thus made up of two systems as follows:

- (i) A radial shear Pa^2/r from O ,
- (ii) A radial shear $\frac{1}{2}PR$ towards C , R being the distance from C of the point considered.

The orthogonal trajectories of the lines of resultant shear are given at once by

$$(Pa^2/r)dr - \frac{1}{2}PRdR = 0$$

or

$$r = C \exp(R^2/4a^2).$$

Since we may add any number of solutions of the above type, it is of interest to examine the result of combining two solutions due to equal and opposite longitudinal shears $Pa^2\pi$ applied along the generators through two points O, O' of the boundary (fig. 1).

In this case the systems of radial shears $\frac{1}{2}PR$ cancel one another, as also do the normal tractions Pz on a cross-section. The shear system is now made up solely of a set of shears Pa^2/r radiating from O and a set of shears Pa^2/r' converging towards O' (r' being the distance from O').

The differential equation of the orthogonal to the lines of shear is thus

$$Pa^2(dr/r - dr'/r') = 0,$$

giving

$$r/r' = \text{const.}$$

Thus the orthogonal trajectories of the lines of shear are the set of coaxial circles having O, O' for real limiting points. The lines of shear themselves are the set of circles passing through O, O' .

If two such circles cut at an angle α , so that our cylinder is in the shape of a cylindrical lens, we have the case of such a lens under equal and opposite longitudinal pulls $Pa^2\alpha$ per unit length applied along the sharp edges, the rest of the curved surface being free from traction.

On the other hand, if we combine two solutions due to equal shears $Pa^2\pi$ applied along the generators through O and O' in the same sense, the system of radial shears $\frac{1}{2}PR$ and the normal tractions Pz are both doubled.

The shear system now consists of

- (i) shears Pa^2/r radiating from O ,
- (ii) shears Pa^2/r' radiating from O' ;
- (iii) shears PR converging to C .

The orthogonal trajectories of the lines of shear are immediately found to be

$$rr' = C \exp(R^2/2a^2).$$

By addition or integration we obtain easily the solution corresponding to any discontinuous or continuous distributions of longitudinal surface tractions, uniform with respect to z .

8—Dealing next with a solid circular cylinder (fig. 2), with a singularity at an internal point O , let O' be the inverse point of O with respect to the circle, and let $(r, \theta), (r', \theta'), (R, \Theta)$ be polar co-ordinates referred to O, O', C as poles, $O'OC$ being taken as the common initial line. Denote OC by c .

Then, along the circumference of the circle

$$A = \frac{1}{2}a^2\Theta + \frac{1}{2}cy = \frac{1}{2}a^2(\theta + \theta') + \frac{1}{2}cy,$$

and

$$\psi_1 = P(\frac{1}{2}a^2\theta - A) = -\frac{1}{2}P(a^2\theta' + cy). \quad (22)$$

Here again, since the function on the right-hand side of (22) is harmonic, it provides the value of ψ_1 at all internal points. The complete values of ϕ and ψ are then

$$\phi = -\frac{1}{2}Pa^2 \log(rr') - \frac{1}{2}Pcx,$$

$$\psi = -\frac{1}{2}Pa^2(\theta + \theta') - \frac{1}{2}Pcy,$$

and the shears in the cross-section consist of

- (i) a set $\frac{1}{2}Pa^2/r$ radiating from O ;
- (ii) a set $\frac{1}{2}Pa^2/r'$ radiating from O' ,
- (iii) a set $\frac{1}{2}PR$ converging to C .

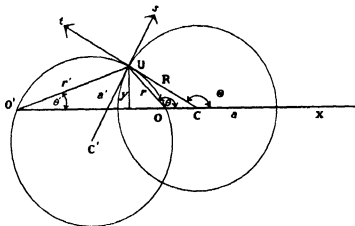


FIG. 2

The orthogonal trajectories of the lines of resultant shear are

$$rr' = C \exp(R^2/2a^2).$$

Any distribution of internal longitudinal pulls may be obtained from the above by summation or integration.

The results of this section and of the last example in the preceding section are summarized together if we consider the radial distributions of shear k/r from O and k/r' from O' (k being any constant) by themselves, and resolve at any point U (fig. 2) along the tangents t, s to the orthogonal circles through U which (i) pass through O, O' , (ii) have O, O' for real limiting points, respectively, the tangents being reckoned positive when along the outwards normal to the orthogonal circle.

$$\text{We then find easily} \quad \widehat{sz} = k/a', \quad \widehat{tz} = k/a,$$

a, a' being the radii of the two circles as shown in fig. 2. It follows at once that suitable radial distributions from the centres of the circles (i) and (ii) annul over these circles \widehat{sz} or \widehat{tz} respectively.

9—So far we have considered only cases in which the cross-section is simply-connected. The case of a multiply-connected cross-section is more complicated and a general treatment will not be attempted.

Limiting our consideration to the case of a cross-section with one hole, i.e. a simple hollow cylinder, let A_1 denote the total area of the hole, A_0 the area of the cross-section of the material, and A_2 the whole area of the section, so that $A_2 = A_0 + A_1$.

Assume for simplicity that there exists inside the material only one singularity O and write as before

$$\phi = -\{(PA_0 \log r)/2\pi\} + \phi_1, \quad \psi = -(PA_0\theta/2\pi) + \psi_1.$$

Over each boundary

$$\psi_1 = P(A_0\theta/2\pi - A) + \text{const}$$

Since, over the inner boundary $Cy(A) = A_1$, $Cy(\theta) = 0$, therefore

$$Cy(\psi_1) = -A_1P.$$

Over the outer boundary $Cy(A) = A_2$, $Cy(\theta) = 2\pi$. Thus

$$Cy(\psi_1) = P(A_0 - A_2) = -A_1P.$$

The cyclic constants of ψ_1 are therefore the same for both boundaries so that $\int (\partial\phi_1/\partial n) ds = 0$ when taken over the complete boundary. It is therefore possible to determine a harmonic function ϕ_1 , which is both acyclic and free from singularities inside, and has the required values of $\partial\phi_1/\partial n$ over the boundary. It is well known that such a determination is unique and it can be obtained, as previously, in terms of the appropriate Green's function H .

We may, on the other hand, use the boundary values of ψ . For, take any convenient point O' inside the hole and write

$$\phi_1 = -\{(PA_1 \log r')/2\pi\} + \phi_2, \quad \psi_1 = -(PA_1\theta'/2\pi) + \psi_2.$$

It now appears that $Cy(\psi_2) = 0$ over both boundaries, and ψ_2 is then one-valued and continuous throughout the region considered, being determined by the condition that

$$\psi_2 - P\{(A_0\theta/2\pi) + (A_1\theta'/2\pi) - A\}$$

is constant over each boundary. This is sufficient to fix ψ_2 uniquely (apart from an additive constant) when the condition that ϕ_2 is acyclic is taken into account.

10—We now proceed to apply these solutions to a few practical problems. We first consider a long rivet or a pile, in the form of a circular cylinder,

gripped tightly all round under uniform lateral pressure Q and driven against friction by uniform normal pressure P_1 applied to one end, against a pressure P_2 applied to the other end. To satisfy the conditions exactly it will be necessary to suppose that, in addition, a distribution of radial shear is applied to both flat ends, but, as this distribution is self-equilibrating, we may assume that, if it is removed, or replaced by any other self-equilibrating distribution, the effects of the change will, by de Saint-Venant's Principle of equipollent loads, become insensible at a distance from the ends.

If γ is the coefficient of friction between the cylinder and the surrounding material, we have two distinct stress-systems to deal with, corresponding to the following surface tractions:

(1) The uniform normal pressure Q all round.

(2) The longitudinal shears γQ due to the friction, balanced by the pressures P_1, P_2 over the ends, where

$$\pi a^2(P_1 - P_2) = 2\pi a l \gamma Q,$$

i.e.

$$P_1 - P_2 = 2l\gamma Q/a,$$

a being the radius of the cylinder and l its length.

The solution corresponding to (1) gives

$$\widehat{xx} = \widehat{yy} = -Q, \quad \widehat{zz} = 0, \quad \widehat{xz} = \widehat{yz} = \widehat{xy} = 0.$$

The solution corresponding to (2) is obtained from (11) by writing $\phi = 0$, $P = (P_1 - P_2)z/l$, and taking for the ends of the cylinder

$$z_1 = -P_1 l / (P_1 - P_2), \quad z_2 = -P_2 l / (P_1 - P_2).$$

This gives, in addition to the longitudinal normal stress

$$\widehat{zz} = (P_1 - P_2)z/l,$$

a radial shear

$$\widehat{rz} = -\frac{1}{2}(P_1 - P_2)r/l.$$

Expressed in cylindrical co-ordinates r, θ, z , referred to the cylinder axis, the only non-vanishing stresses are

$$\begin{aligned} \widehat{rr} = \widehat{\theta\theta} &= -Q, \quad \widehat{zz} = (P_1 - P_2)z/l, \quad \widehat{rz} = -\frac{1}{2}(P_1 - P_2)r/l, \\ &= 2\gamma Qz/a, \quad = -\gamma Qr/a. \end{aligned}$$

The lines of principal stress in a meridian plane are given by the equation

$$\tan 2\chi = 2\widehat{rz}/(\widehat{zz} - \widehat{rr}) = -r/(z - z_0),$$

where $z_0 = -a/2\gamma$ and χ is the inclination of a line of principal stress to the axis of z .

Refer to polar co-ordinates (R, Θ) with the point $J(r = 0, z = z_0)$ as pole and the positive axis of z as initial line, so that $r/(z - z_0) = \tan \Theta$. We have

$$\chi = -\Theta/2.$$

The tangent to a line of principal stress makes an angle $3\Theta/2$ with the radius vector, whence (paying attention to signs) the differential equation of a line of principal stress is

$$(dR/R) + \cot(3\Theta/2) d\Theta = 0,$$

leading to

$$R^3 \sin(3\Theta/2) = \text{const.}$$

These lines are sketched in fig. 3. It is interesting to note that they are

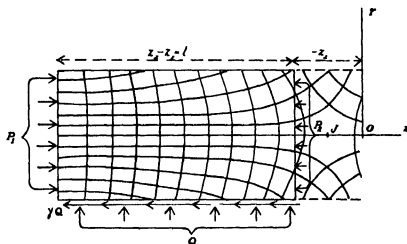


FIG. 3

independent, as regards shape and size, of the pressures P_1 , P_2 and Q , and of the coefficient of friction γ .

Their position, however, depends on that of the point J . This point is an "isotropic point", at which the three principal stresses are equal.

The condition that the isotropic point shall fall inside the cylinder is that $z_0 < z_2$, or $Q > P_2$.

In the case of a peg or rivet driven through a long hole, when we can take $P_2 = 0$, the above condition is always satisfied, so that an isotropic point exists, but, unless γ is small, this is usually within the region affected by the end-conditions, and these, as we have seen, will not as a rule be exactly satisfied.

In the case of a pile driven into the ground, P_2 will generally be considerable, and the isotropic point will be outside the cylinder. This is the

case actually represented in fig. 3, but the lines of principal stress have been drawn in the prolongation of the cylinder up to $z = 0$, so that the other cases are readily supplied by imagining the right-hand boundary in fig. 3 displaced in the direction of z through any required distance.

The magnitudes of the principal stresses are easily found to be

$$(\gamma Q/a)(z + z_0 + R), \quad (\gamma Q/a)(z + z_0 - R), \quad (\gamma Q/a)(2z_0),$$

the first two being in the meridian plane.

The stress differences are then

$$(\gamma Q/a)(z - z_0 + R), \quad (\gamma Q/a)(z - z_0 - R), \quad (\gamma Q/a)(2R),$$

and since R is always greater than $|z - z_0|$, the numerically greatest stress difference is always the third, namely the stress difference in the meridian plane.

Taking this greatest stress difference as the criterion for failure, it appears that the points where yielding is first to be expected are the points most distant from J , and these in general will be on the perimeter of the end $z = z_1$, where the driving pressure P_1 is applied.

In practice the condition that the lateral pressure Q is constant throughout the length of the cylinder will rarely be realized. The mode of its variation must necessarily depend on the nature of the surrounding material. In the case of a rivet fitting tightly a hole in elastic material, the lateral expansion of the rivet has to be taken into account. This is given by equations (18) as

$$2\mu(u_r)_{r=a} = -\lambda(P_1 - P_2)(az/l)/(3\lambda + 2\mu)$$

+ a constant part due to lateral pressures Q .

Thus $(u_r)_{r=a}$ increases as z increases negatively and the fit is tighter near the "hammer" end, for this reason. This must increase the lateral pressure.

An exact solution when Q varies is still to be found, but, if Q varies slowly according to a known law, an approximation can be obtained by assuming the present form of solution to hold good for every element of the cylinder, in much the same way as is done in the ordinary theory of continuous beams where solutions valid only for constant bending moment or for constant shear are applied to cases where these quantities are variable.

11—The second problem to be considered will be similar to the first, except that the normal traction $(\hat{r}\hat{r})_{r=a}$ over the curved boundary, instead of being uniform, will be assumed to obey the law

$$(\hat{r}\hat{r})_{r=a} = -Q(1 - \cos n\theta),$$

and, in consequence, $(\hat{r}\bar{z})_{r=a} = -\gamma Q(1 - \cos n\theta)$.

To simplify, we take $P_2 = 0$ so that $z_2 = 0$, $z_1 = -l$, $P = P_1/l$.

The part of the solution due to the shears $\hat{r}\bar{z}$ is obtained by assuming $\phi = Ar^n \cos n\theta$. Then

$$\hat{r}\bar{z} = -\frac{\partial\phi}{\partial r} - \frac{1}{2}Pr,$$

and, when $r = a$, we have

$$-nAa^{n-1} = \gamma Q, \quad P = \frac{2\gamma}{a} \times Q,$$

so that $P_1 = 2\gamma l Q/a$, as before.

We have accordingly, from this part,

$$\hat{r}\bar{z} = -(\gamma Qr/a) + \gamma Q(r/a)^{n-1} \cos n\theta, \quad \hat{z}\bar{z} = 2\gamma Q(z/a).$$

The part of the solution due to the normal tractions $\hat{r}\bar{r}$ can again be broken up into two. The first is due to $(\hat{r}\bar{r})_{r=a} = -Q$ and gives $\hat{r}\bar{r} = \hat{\theta}\bar{\theta} = -Q$, as before.

To obtain the second, remembering that the cylinder must be long, we use the solution for plane strain. The stresses are then derived from a stress-function χ by the equations

$$\hat{r}\bar{r} = (1/r^2)(\partial^2\chi/\partial\theta^2) + (1/r)(\partial\chi/\partial r), \quad \hat{r}\bar{\theta} = -(\partial/\partial r)(\partial\chi/r\partial\theta), \quad \hat{\theta}\bar{\theta} = \partial^2\chi/\partial r^2.$$

χ satisfies the biharmonic equation, and we can assume

$$\chi = (Ar^n + Br^{n+2}) \cos n\theta.$$

A and B are then easily found from the boundary conditions

$$A = -\frac{1}{2}(Qa^{-n+2})/(n-1), \quad B = \frac{1}{2}(Qa^{-n})/(n+1),$$

whence $\hat{r}\bar{r} = \frac{1}{2}Q\{n(r/a)^{n-2} - (n-2)(r/a)^n\} \cos n\theta$,

$$\hat{\theta}\bar{\theta} = \frac{1}{2}Q\{-n(r/a)^{n-2} + (n+2)(r/a)^n\} \cos n\theta,$$

$$\hat{r}\bar{\theta} = \frac{1}{2}nQ\{(r/a)^n - (r/a)^{n-2}\} \sin n\theta.$$

This plane strain solution introduces a stress $\hat{z}\bar{z}$ given by

$$\hat{z}\bar{z} = \eta(\hat{r}\bar{r} + \hat{\theta}\bar{\theta}) = 2\eta Q(r/a)^n \cos n\theta,$$

where η is Poisson's ratio, as before.

Provided $n > 1$, this stress forms a self-equilibrating system over the ends and will not affect total end-conditions.

Thus, in the body of the cylinder, the stresses are given by the values of $\hat{r}\bar{z}$, $\hat{r}\bar{r}$, $\hat{\theta}\bar{\theta}$, $\hat{r}\bar{\theta}$ already obtained and by

$$\hat{z}\bar{z} = 2\gamma Q(z/a) + 2\eta Q(r/a)^n \cos n\theta.$$

The above problem provides a rough representation of an elongated rifle bullet (the effect of the pointed end being neglected) forced through a barrel with n grooves and lands, the spiral shape of the rifling being neglected.

A hollow shell could be treated in like manner, there being no traction across the inner boundary. Here, in the part due to the shears $\widehat{r_z}$, we assume $\phi = C \log r + Ar^n \cos n\theta + A'r^{-n} \cos n\theta$. The solution for the normal tractions $\widehat{r_r}$ is known (see Coker and Filon 1931, § 4.31). The results are more complicated than those for the solid cylinder; but present no special analytical difficulty.

12—The third problem is that of a long cylindrical roller, tightly gripped between two rough parallel planes and forced through against friction, in the direction of its axis, by a thrust P_1 applied to the end $z = -l$, the end $z = 0$ being free from thrust.

Let N be the normal pressure per unit length, supposed to be uniformly distributed, and $F = \gamma N$ the friction, also per unit length. We have

$$P_1 = 2\gamma Nl = 2Fl.$$

As before, we treat separately the stress systems due to F and N .

The solution due to the concentrated shears F is immediately written down from the last example of § 7, where $P = F/\pi a^2$; C , O and O' being the centre of the cross-section and the points of contact with the parallel planes, respectively. Owing to symmetry the system of shears $\widehat{x_z}$, $\widehat{y_z}$ form a self-equilibrating system.

The solution due to the normal pressures N is given by Love (1927, Art. 155). The stress-system in the plane of the cross-section consists of

- (i) A uniform all-round tension

$$\widehat{r_r} = \widehat{\theta\theta} = N/\pi a,$$

- (ii) Radial distributions of pressure, of amounts

$$\widehat{r_1 r_1} = -(2N \cos \theta_1)/\pi r_1, \quad \widehat{r_2 r_2} = -(2N \cos \theta_2)/\pi r_2,$$

with centres O , O' respectively, (r_1, θ_1) being polar co-ordinates with pole O and initial line OO' , (r_2, θ_2) polar co-ordinates with pole O' and initial line $O'O$.

This stress system, if associated with plane strain, involves an axial normal stress $\widehat{z z}$ given by

$$\eta(\widehat{r_1 r_1} + \widehat{r_2 r_2}) + \gamma N/\pi a.$$

It is easy to show that the resultant of the stresses

$$\widehat{z z} = \eta(\widehat{r_1 r_1} + \widehat{r_2 r_2})$$

is in this case an axial thrust $4\eta Na$ acting through C . If we now add the solution due to a uniform tension

$$\widehat{z}z = 3\eta N/\pi a,$$

we have left, over and above the thrust P_1 on the end $z = -l$, a self-equilibrating system of tractions over each end, and the problem, at a distance from the ends, may be considered as solved.

13—The fourth problem will be similar to the last one, except that both end-sections are free from thrust and the two planes are pushed in opposite senses parallel to the axis of the roller, so that the frictions F have opposite signs. We still have (if the friction is limiting)

$$\gamma N = F = \pi a^2 P.$$

The solution due to the normal pressures N remains as before, a uniform tension $\widehat{z}z = 3\eta N/\pi a$ being superimposed to make the normal tractions over an end self-equilibrating.

The solution due to the forces F is written down from the second example of § 7. In this case, however, the shears over an end do not form a self-equilibrating system, but have a resultant $2aF$ along the diameter OO' . This is inevitable from the conditions of statical equilibrium, since the couple $2alF$ of the frictions must be balanced by forces on the ends.

If the resultant of the shears were removed, a readjustment of the normal pressures N would follow, which would provide the necessary balancing couple. The conditions required by the type of solution we have been considering throughout would no longer be satisfied, and a far more complicated problem would result.

In practice it would seem reasonable to assume that, if the roller is long, uniform conditions are established away from the ends. This involves that the necessary lateral thrusts are introduced near the ends and are transmitted to the more central portions of the roller by means of shears in the cross-section, approximating more and more closely to the distribution given by the solution above.

SUMMARY

A stress system for which the stresses $\widehat{x}x$, $\widehat{y}y$, $\widehat{z}z$, in the plane of xy , all vanish may be referred to briefly as "antiplane" stress. Such a type of stress includes simple tension parallel to z , flexure under constant bending moment, and de Saint-Venant's solutions for torsion and flexure.

It is shown that it also includes a type of solution which appears to have been hitherto overlooked, in which the stresses reduce to

$$\bar{xz} = -\partial\phi/\partial x - \frac{1}{2}Px, \quad \bar{yz} = -\partial\phi/\partial y - \frac{1}{2}Py, \quad \bar{zz} = Pz,$$

ϕ being a harmonic function.

The characteristics of this solution are investigated in the case of a cylinder whose axis is parallel to the axis of z . It is proved to involve internal pulls along axial fibres, or longitudinal shears along the curved surface of the cylinder.

The function ϕ is determined in simple cases and four examples are given where, by combination with known systems of plane strain or plane stress, such an antiplane stress provides the solution of certain practical problems, in which a long circular cylinder is forced through axially against the friction of constraints.

REFERENCES

- Coker, E. G. and Filon, L. N. G. 1931 "Photo-Elasticity." Cambridge.
 Love, A. E. H. 1927 "A Treatise on the Mathematical Theory of Elasticity."
 Cambridge.
-

On the Action of Ultra-Violet Sunlight upon the Upper Atmosphere

By M. N. SAHA, F.R.S., *Allahabad University*

(Received 19 August 1936)

I—INTRODUCTION

The ordinary solar spectrum extends, as is well known, to about $\lambda 2913$, the more ultra-violet parts being cut off by ozone absorption in the upper atmosphere. We have thus no direct knowledge of the distribution of intensity in the solar spectrum beyond $\lambda 2913$, as it will appear to an observer situated outside the atmosphere of the earth. But it is now recognized that a number of physical phenomena is directly caused by the photochemical action of this part of sunlight on the constituents of the upper atmosphere. Such phenomena are (1) the luminous spectrum of the night sky and of the sunlit aurora,* (2) the ionization in the E, F and other layers which is now being intensely studied by radio-researchers all over the world,† (3) the formation and equilibrium of ozone (see Ladenburg 1935), (4) magnetic storms and generally the electrical state of the atmosphere.

Formerly it was a debatable point whether some of these phenomena were not to be ascribed to the action of streams of charged particles emanating from the sun. There seems to be no doubt that the polar aurora and certain classes of magnetic storms are to be ascribed to the bombardment of molecules of N_2 and O_2 by such charged particles, for these phenomena show a period which is identical with the eleven year period of the sun, and are found in greater abundance, the nearer we approach the magnetic poles.‡ But there now exists no doubt that the ionization observed by means of radio-methods in the E and F_1 regions, their variation throughout day and night, and at different seasons is due to the action of ultra-violet sunlight. This was decisively proved by observations during several total solar eclipses since 1932 (Appleton and Chapman 1935). The luminous night-sky spectrum, though it has certain points of similarity to the polar

* For a general account of the spectrum of the luminous night sky, see Dejaridin (1936).

† For general information regarding investigation on the ionosphere, see Mitra and others (1936) and Appleton (1936).

‡ See for general information article by Störmer (1931). The frequency of auroræ appears to reach a maximum 20° from the magnetic pole.

aurora, is on the whole widely different, and is found on nights free from electrical disturbances. The prevailing opinion is that it is mainly due to the ultra-violet solar rays, i.e. in the course of the day sunlight is stored up by absorption by the molecules in the upper atmosphere, and again given up during the night, in one or several steps, as a fluorescence spectrum. According to S. Chapman (1930) the formation of the ozone layer and its equilibrium under different seasonal conditions is also to be mainly ascribed to the action of ultra-violet sunlight. In the following paper an attempt will be made to discuss some of these questions in as rigorous a way as is possible with our present knowledge. It is evident that an adequate discussion is possible only if we have a good knowledge of (1) the distribution of intensity in the solar spectrum beyond $\lambda 2900$, (2) the photochemical action of light of shorter wave-length than $\lambda 2900$ on the constituent molecules of the upper atmosphere, which are mainly oxygen and nitrogen. We shall first consider (1).

2—THE ULTRA-VIOLET SPECTRUM OF THE SUN

The disappearance of sunlight below $\lambda 2900$ has long been known, through the researches of Fowler and Strutt (1917), and of Fabry and Buisson (1913) and others, to be due to the absorbing action of a layer of O_3 (equivalent in amount to a 3 mm. column at N.T.P.) formed in the upper atmosphere. The long series of works by Götz, Meetham and Dobson (1934) and Regener (1934) has shown that this layer extends from about 20 km., reaches a maximum concentration at 30 km., and probably does not extend much beyond 50 km. But, as we shall see presently, the principal photochemical reactions in N_2 and O_2 , including ionization, are produced only by light of wave-length $< 3000 \text{ \AA}$, and hence in order that a correct estimate may be made of the action of sunlight on O_2 and N_2 gas above the ozone layer, it is necessary for us to have a detailed knowledge of the emission of the sun below $\lambda 3000 \text{ \AA}$. In the absence of direct knowledge, recourse is made to extrapolation, i.e. emission below $\lambda 3000 \text{ \AA}$ is supposed to be identical with that of a black body at 6800°K . The justification for such an assumption is found in the work of H. H. Plaskett* and Fabry that the intensity distribution in the solar spectrum in regions free from any absorption lines agrees very well with that of a black body at 6800°K . But this evidence is at best indirect, and many observers hold that solar emission in the ultra-

* For detailed report on this very intricate problem, and generally on the temperature of stars, see Brill (1932); for later work, see the same, *Ergänzungsband*, 7, but usually discussion is confined only to available radiation from stars.

violet is subject to a good deal of fluctuation and differs widely from that of a black body.*

3—THE NEGATIVE BANDS OF NITROGEN IN THE SPECTRUM OF THE NIGHT SKY

It is now possible to produce further evidence that ultra-violet emission from the sun is widely different from black-body radiation, but we shall at the outset discuss in full only *one line of evidence which does not appear to admit of any other interpretation*. This is the appearance of the earlier members of the first negative bands of nitrogen in the spectrum of the upper atmosphere when illuminated by sunlight. These bands have the wave-lengths $\{(0,0)—3914\}$, and $\{(0,1)—4278\}$, and on unimpeachable spectroscopic grounds they have been ascribed to N_2^+ , *vide* § 6.†

These bands are very intense in the spectrum of the polar aurora, where they are produced by the bombardment of N_2 gas by electrons. The intensity is found to be the same as that of the green line of O, $\lambda 5577.35$. But the lines are also found in the spectrum of the night sky, though so faintly that some observers have even denied their existence in the night-sky spectrum. But

* The opinion that the ultra-violet radiation from the sun differs widely from that of a black body at 6800° K. has been expressed by many workers, from different points of view. Petit (1935) finds, from actual measurements between $\lambda\lambda 4100-3000$, that the emission curve shows no resemblance to that of a black body.

Mulders (1935) finds that the solar emission curve (when corrected for absorption) shows a form which bears no resemblance to a black-body curve. From $\lambda 2500$ to $\lambda 4100$ it approximates a black-body curve at 7140° K., but between $\lambda\lambda 4000-3000$ it approximates a black body curve at 4800° K. These determinations are subject to a good deal of uncertainty, as no account is taken of absorption by O_2 and N_2 in the uppermost layers, as described in the present paper, and therefore they do not apply to the region discussed in this paper.

Speculations below $\lambda 3000$. To explain terrestrial magnetic storms and aurorae, Mars and Hulburt (1929) proposed a hypothesis that the sun may, from time to time, be subjected to outbursts of ultra-violet emission localized at certain points. These flares of radiation may have 10^6 times the intensity of black-body radiation between $\lambda\lambda 500-1000$, according to the above-mentioned authors.

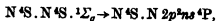
Gurney (1928) expresses the view that the solar radiation in the ultra-violet (near $\lambda 584.4$) is in excess of black-body radiation by a factor of 10^6 . In a note (Saha 1936) the present author pointed out that the flash of bands due to CO^+ and N_2^+ in the tail of comets can be explained if we suppose that the sun shoots out high energy photons and ionizes CO and N_2 in the tail in the way proposed in § 3 of this paper. In the course of a discussion, Professor H. N. Russell pointed out that once N_2 and CO are ionized, they can be maintained in excited states of N_2^+ and CO^+ by visible radiation from the sun. The subject of cometary spectra will be taken up in a subsequent paper.

† For photographs of N_2^+ bands in the aurora, see Slipher and Sommer (1929); see also Slipher (1933).

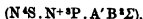
the issue has been cleared up by the following observation due to Slipher (1933):

"But the negative nitrogen bands typical of auroral display spectra do not accompany this chief yellow line (green) in the sky, except when auroral displays are actually present. However, these nitrogen bands, I found, could be photographed in the morning and evening skies if brief exposure were made at the moments, when the last and first traces of sunlight touch the high atmosphere. Thus the day, as it were, begins and ends with a sort of auroral flash."

The physical meaning of this observation becomes clear if we consider the genesis of the N_2^+ bands. For this purpose, it is necessary to turn to fig. 1, which is the energy-level diagram of N_2 and N_2^+ . We notice that the first negative bands are due to transition of the electron from a level called A' ($N^4S, N^+^3P, B^3\Sigma$)* to the normal level of N_2^+ called X' ($N^2D, N^+^3P, ^3\Sigma_g^+$). The energies of excitation of these two levels have been obtained from both cathode-ray bombardment and from spectroscopic analysis, both of which yield concordant values. As a glance at fig. 1 shows, the first ionization potential (excitation of the X' level) is 15.55 V, and the second ionization potential (excitation of A level) is 18.68 V. But Slipher's observation shows that the A' level of N_2^+ is produced by sunlight acting directly on the N_2 molecules of the upper atmosphere. It therefore becomes necessary to find out how these bands are produced by direct photochemical action of sunlight, without the intervention of any bombarding electron. This is furnished by the study of the absorption spectrum of N_2 by Hopfield (1930), who found a number of absorption bands at $\lambda\lambda 723, 694, 681, 675, 671$ which pass to a continuous absorption at $\lambda 660$. Mulliken (1933) shows that the most probable explanation of the origin of these bands is as follows:



i.e. one of the component atoms remains unchanged. In the other atom, one electron passes to the ns state, thus giving rise to a number of 4P terms in a Rydberg series. The limit corresponds to the state of N_2^+



which is marked by a continuous absorption at $\lambda 660$, which corresponds to the ionization of N_2 to the A level of N_2^+ . What is particularly important to notice is the fact that Hopfield did not find any trace of continuous absorption at the first ionization potential of N_2 , corresponding to the X' level.

Several investigators (for example, see Price and Collins 1930, particularly

* This notation means that one N atom is in a 4S state, the other is singly ionized and is in a 3P state and the molecule as a whole is in a $^3\Sigma$ state, and so on.

para. 2, p. 715) have subsequently found that many diatomic molecules show a similar type of strong absorption leading to one or more higher ionization potentials, *the absorption leading to the lowest ionization potential being very feeble or sometimes totally absent.*

Hopfield's experiment therefore proves that light of wave-length $\lambda 794$ cannot directly ionize N_2 to the N_2^+ normal state, i.e. a quantum having an energy content of 15.55 V (X' level), $\lambda 794$, has no direct action on N_2 , though the energy is sufficient for raising N_2 to the lowest state of N_2^+ . But a quantum of energy content of 18.68 V, $\lambda 661$, can directly ionize N_2 to the A level. This supplies the clue to the interpretation of Slipher's result. Sunlight of wave-length $< \lambda 661$ produces direct ionization of N_2 to the A state of N_2^+ , which, being excited, emits the negative bands and reverts to the normal N_2^+ state (X). This explains the flash of negative bands at sunrise and sunset. After the withdrawal of sunlight, the excited N_2^+ ions speedily revert to the normal state, and we are left only with N_2^+ normal ions. These are incapable of radiating, and in the course of the night may be neutralized by the direct capture of an electron in different excited states of N_2 , but it is quite probable that when the colliding electrons have sufficient velocity, a number of N_2^+ (X') ions may be further excited to the A state. This process may account for the feeble emission of negative bands during the night observed by Sommer and certain other workers.*

4—IONIZATION OF THE NITROGEN MOLECULE IN THE UPPER ATMOSPHERE

The above arguments therefore prove in a very conclusive manner that N_2 is ionized above a height of 200 km. directly to N_2^+ by the direct action

* The author has discussed the line of argument followed here with many workers on the field in Europe and America, in the course of his travels during 1936, but the only serious objection against it was raised in the course of a friendly conversation with Dr Wulf, of the Bureau of Soils and Agricultural Research, Washington, D.C. Dr Wulf thinks that the ionization of N_2 to N_2^+ is due to the photons having the wave-length $\lambda 794$, corresponding to the first ionization potential of 15.55 V; the N_2^+ ions produced in this way further absorb the quanta representing the second negative bands of N_2 , and thus raise N_2^+ to the N_2^+ excited state A'. The flash of negative bands is due to fluorescence of these excited states of N_2^+ .

The mechanism of N_2^+ luminescence postulated by Dr Wulf appears unlikely, as according to the experiments of Hopfield, N_2 shows no, or extremely feeble, absorption at $\lambda 794$. If we suppose that the absorption exists at all, it must be very small, hence we should have to suppose that these rays penetrate to very low depths of the atmosphere. The observational fact that N_2^+ bands are obtained at heights exceeding 200 km., and not below, shows that the photons giving rise to N_2^+ luminescence are absorbed completely by a very small amount of N_2 . This can apply only to continuous absorption at $\lambda 660$, corresponding to the second ionization potential.

Regarding emission of N_2^+ bands by the night sky, free from aurora, evidence appears to be positive, Dejardin (1936, p. 10). The bands are extremely faint.

of sunlight. It is difficult to estimate the number of N_2^+ excited molecules thus produced, without photometric measurement of the intensity of the bands, and other relevant laboratory experiments. But we obtain an idea of the number of free electrons in this region from measurement of the F_2 layer ionization. This is estimated by Appleton (1936) and others to be nearly 5×10^5 electrons per c.c., and we can assume that a fair proportion of these electrons is produced by the direct ionization of N_2 to N_2^+ by the action of ultra-violet sunlight. Apart from this evidence, the very intensity of N_2^+ bands during daylight flash, and their intensification in a sunlit aurora (Störmer 1931), is definite proof that sunlight produces considerable ionization of N_2 to N_2^+ . It will now be shown that if the radiation from the sun be supposed to be the same as that given by a black body at 6500° , the number of quanta available having an energy content greater than 18.68 V (excitation potential of N_2 to the A state of N_2^+) is hopelessly inadequate for the purpose. The argument is as follows:*

The number of quanta of frequency greater than ν , and falling normally on unit surface of the earth per sec., is given by

$$N_\nu = 3 \times 10^{17} I_1,$$

where

$$I_1 = \int_{x_0}^{\infty} \frac{x^2}{e^x - 1} dx \quad \text{and} \quad x = \frac{h\nu}{KT}. \quad (1)$$

For $\lambda = 660 \text{ \AA}$ corresponding to 18.68 e-volts, $I_1 = 3 \times 10^{-13}$, hence we have

$$N_\nu = 10^4. \quad (1')$$

So the solar rays can produce only 10^4 ions of N_2^+ (A state)/sec./cm.² in the whole depth of the atmosphere. But according to a calculation by Chapman (1931), the total number of ions to be produced per sec. for maintaining the *total ionization* is 3×10^{10} . Of course, the total ionization is not entirely due to N_2^+ , probably the greater part is due to ionization of O_2^+ . But the intensity of N_2^+ bands shows that at least a substantial part, say one-tenth, is due to ionization of N_2 . Hence we conclude that the solar ultra-violet light of wave-length $< \lambda 660$ is about a million times more intense than that given by a black body at a temperature of 6500°K . This conclusion, *to which it is difficult to see any alternative*, brings out the necessity of investigating the ultra-violet emission of the sun with greater care than has hitherto been done. From arguments which are given elsewhere, it appears probable that if we could observe the solar spectrum outside the atmosphere of the earth, it would appear very much like those of planetary

* For further details about this calculation see Saha (1935).

nebulae, i.e. composed of a faint continuous background superimposed with bright emission lines of H, He and He^+ , Fe^+ , Fe^{++} , and other elements which are abundant in the atmosphere of the sun, and which have their resonance lines in this part of the spectrum.

Bearing these points in mind, we now turn to a critical examination of the results on the night-sky spectrum and on the ionosphere.

5—DISTRIBUTION OF ELEMENTS IN THE UPPER ATMOSPHERE

A rigorous theoretical treatment of all the upper air phenomena requires a knowledge of the distribution of elements at different heights of the atmosphere, and its variation by night and day. We shall illustrate the point by an example.

In the foregoing paragraphs, we discussed in detail the appearance of N_2^+ bands as a flash in the upper atmosphere produced by the disappearing light of the evening sky, or by the dawning light of the morning sky. Rough estimates show that the phenomenon occurs at a height of 200 km. (see Dejardin 1936, p. 10). Why does not this flash extend below? The obvious answer is that at this height solar light of wave-length $< \lambda 660$ is completely absorbed by the layer of gas traversed. We have next to link up this phenomenon with the laboratory experiments dealing with the intensity of absorption of N_2 gas at wave-lengths below $\lambda 660$. Supposing we find, from accurate laboratory experiments, that 1/500 cm. of N_2 at N.T.P. reduces the intensity of $\lambda 660$ to 1/eth of its value, and suppose we assume that the action of light is inappreciable when the intensity falls to 10^{-3} of its primitive value. This can be achieved by $x \times 1/500$ cm. of N_2 gas at N.T.P., where $e^x = 10^3$, i.e. $x = 7$, and the thickness of gas is 0.014 cm. at N.T.P., i.e. we conclude that the amount of N_2 gas traversed by sunlight at nearly horizontal incidence at a height of 200 km. is 0.014 cm. of N_2 at N.T.P., and hence there will be no ionization below 200 km.

At the present time our knowledge of the absorption coefficient of every ray below $\lambda 3000$ in O_2 and N_2 is in an extremely unsatisfactory state, but probably this will be available before long. Knowledge of this absorption coefficient will enable us to estimate quantitatively the photochemical reactions produced in the upper atmosphere, provided we have a knowledge of the distribution of the elements at different heights. We should therefore have ready for use a table which tells us of the quantity of gas of each kind (N_2 or O_2) which lies above a height z . It is usual to express this quantity in lengths, i.e. centimetres of the gas at N.T.P., 1 cm. being equal to 2.79×10^{19} mol. On the surface of the earth the quantity is given by H ,

the height of the homogeneous atmosphere. At any height z , let the corresponding quantity for a gas x be given by H_x^z . Then we have $H_x^z = P_x^z / \rho_0^x$, where P_x^z is the partial pressure of the gas x at height z , and ρ_0^x is the density of the gas at N.T.P. To calculate P_x^z we take the hydrodynamical equation

$$dp = -\rho g dz. \quad (2)$$

We have omitted z and x from p and ρ for the sake of elegance. Now we have

$$p = \frac{RT}{M} \rho,$$

where M = molecular weight, R = universal gas content. Hence

$$\frac{dp}{p} = -\frac{Mg dz}{R T} \quad \text{or} \quad \log_e \frac{P_x^z}{P_0^x} = -\frac{gM_x}{R} \int_0^z \frac{dz}{T}. \quad (3)$$

The integral $\int_0^z \frac{dz}{T}$ is common for all gases. Let it be denoted by β . Then we have

$$P_x^z = P_0^x e^{-\frac{gM_x}{R}\beta}, \quad \text{and} \quad H_x^z = H_0^x e^{-\frac{gM_x}{R}\beta}, \quad (4)$$

where $H_0^x = \frac{P_0^x}{\rho_0^x}$ denotes the quantity of gas at N.T.P. which is equivalent to the whole column in the atmosphere.

An approximate (not accurate, but sufficient for the height considered by us) value of the integral is obtained if we know the total pressure at z , for considering that the mixture does not change much, we have

$$\log_e \frac{P}{P_0} = -\frac{gM}{R} \int_0^z \frac{dz}{T} = -\frac{gM}{R} \beta, \quad (5)$$

or

$$\beta = \frac{R}{gM} \log_e \frac{P}{P_0},$$

where P, P_0 are total pressures, and M = mean molecular weight. We have

$$H_x^z = H_0^x \left(\frac{P}{P_0} \right)^{\frac{M_x}{M}}. \quad (6)$$

The calculation of the integral $\int_0^z \frac{dz}{T}$ is one of the main problems in Meteorology, but as it cannot be obtained, we have taken empirical values of (P/P_0) as obtained directly or in the ionosphere by radio experiments, and tried to calculate H_x^z . The procedure cannot be far wrong.

TABLE I

Height in km.	Pressure in mm. of Hg	Amount of gas in metres	
		N ₂	O ₂
0	760	6445	1508
10.5	215	1724	342
20	42	386	60
30	10	94.8	12
40	2	20	2
50	10 ⁻¹	110 cm.	7 × 10 ⁻⁴ = 7 cm.
60	10 ⁻²	11 cm.	0.57 cm.
100 E	10 ⁻³	—	0.044 cm.
140 E'	—	—	—
180 F ₁	—	—	—
250 F ₂	10 ⁻⁴	14 × 10 ⁻³ cm.	—

6—A GENERAL REVIEW OF THE SPECTRA OF NITROGEN AND OXYGEN

Recent work on spectra of the night sky and sunlit aurora prove that even up to 200–300 km., where these phenomena appear, the atmosphere consists chiefly of O₂ and N₂. Helium and hydrogen, which have sometimes been postulated to exist, still require confirmation. Hydrogen appears to be definitely absent. Paneth has recently devised methods for estimating quantitatively amounts of helium at different heights.

The recent state of our knowledge of the spectrum of N₂ and N has been very fully described in a report by L. S. Mathur and P. K. Sengupta (1936),* and as this knowledge is important for the discussion which follows, we give a brief account of the spectra of these elements with some necessary additions, chiefly on absorption. Further, the energy-level diagram, reproduced below, will make perusal of this subject somewhat easier.

The night-sky spectrum shows the following bands due to N₂ and N₂⁺:

- (i) Vegard-Kaplan intercombination bands.
- (ii) The first positive bands.
- (iii) The second positive bands.
- (iv) The first negative bands due to N₂⁺.

The origin of these bands is illustrated in the energy-level diagram, fig. 1.

The figures on the *y*-axis denote the excitation potential of the different levels in V. The symbols ⁴S + ¹D, etc., on the sides of the *y*-axis show the state of the atoms composing the molecule which give rise to the particular level. Thus (¹D + ³P) C³Π_u denote that the component N atoms of the molecule are in the ¹D and ³P states, and they give rise to a ³Π_u term,

* For a general report on the spectra of O₂ and N₂ see Sponer (1935).

* Reprinted from *Proc. Nat. Inst. Sci. Ind.* 1, 232, 1935.

(a) The Birge-Hopfield bands $N_2(^4S+^4S)^1\Sigma \rightarrow N_2^1\Pi_u$ bands. The vibration formulae for these bands are given by

$$\nu = 68962.7 + (1678.96\nu' - 13.318\nu'^2 - 0.0354\nu'^3) - (2345.16\nu'' - 13.93\nu''^2).$$

These bands extend from $\lambda 1450$ to $\lambda 1228$, corresponding to $\nu'' = 0$. No quantitative estimate is available of the intensity of this absorption. It is stated by Lyman (1928) that 9 mm. of air at N.T.P. is sufficient to show this absorption. The bands were first obtained in absorption, and identified by Sponer.

The corresponding amount is attained at a height of about 60 km., but it is probably unnecessary to consider this absorption, as oxygen has far stronger absorption in this region. In fact, Ladenburg finds that 0.002 cm. of O_2 at N.T.P. reduces $\lambda 1450$ to one-half its strength, so it is clear that this radiation is almost completely absorbed above the E layer by oxygen molecules.

(b) The Vegard-Kaplan bands ($X'\Sigma_g \rightarrow A^3\Sigma_g^+$). These bands are given by the formula

$$\nu = 49774.4 + (1446.46\nu' - 13.93\nu'^2) - (2345.46\nu'' - 14.45\nu''^2),$$

and absorption by normal molecules should extend from $\lambda 2000$ towards the ultra-violet. But as the values of r for the normal and excited states are widely different ($r' = 1.29$ cm., $r'' = 1.09$ cm.), transitions from $\nu'' = 0$ to some high value of ν' will be more probable, according to the Franck-Condon Principle. Such bands will occur at about $\lambda 1700$.

Though no laboratory experiment has yet been done with the express intention of obtaining the absorption of Vegard-Kaplan bands, it appears that they were recorded by Hopfield (1928) some years ago. He found that with 0.06 to 19 m. of N_2 column at N.T.P., the following bands were obtained: $\lambda\lambda 1742.4, 1728.4, 1701.4, 1688.3, 1666.3, 1650.2$. These are possibly Vegard-Kaplan absorption bands.

These wave-lengths are also strongly absorbed by O_2 , but not so strongly as $\lambda 1460$. Hence it is quite plausible that by means of this absorption, a large proportion of N_2 molecules is excited to the $N_2(N^4S+N^4D)A^3\Sigma_u^+$ stage, which is the final level of the first positive bands of N_2 . It therefore appears plausible that, during daytime, a large proportion of N_2 molecules will be found in the excited state of $N_2A^3\Sigma_u^+$, and molecules in such excited states will give rise to absorption of the first positive bands. But it is not possible to make any calculation of the proportion of such excited molecules, as all the data required for the purpose are lacking. These are (1) the intensity of solar light in the region $\lambda\lambda 2000-1700$; (2) the intensity of

absorption of these bands in nitrogen gas. But the process is probably confined to over 30 km., as at this height the amount of N_2 gas is about 100 m. at N.T.P., which is the amount required for complete absorption of these bands.

As purely a tentative measure, an attempt has been made to find out whether any of the well-known N_2 band lines belonging to the first or the second positive bands, or due to N_2^+ , occur amongst the unidentified lines of the telluric part of the Fraunhofer spectrum. The results, which are not very conclusive, will be communicated in a later paper.

THE NEGATIVE BANDS OF NITROGEN

No estimate is yet available of the intensity of continuous absorption of N_2 at $\lambda 661$ which is responsible for ionizing N_2 to the N_2^+ A' state and thus giving rise to the negative bands of N_2^+ observed in the first sunlight flash of morning or evening hours. But it appears that the N_2^+ ion is a normal constituent of the upper atmosphere during daytime, and if there be a sufficient number of them, we may obtain the strongest N_2^+ bands in the Fraunhofer spectrum as telluric lines. The relevant data will be discussed elsewhere. They are also not very conclusive.

Can nitrogen be photochemically decomposed into atoms by sunlight?

It is well known that in the case of most diatomic molecules, and particularly in the case of O_2 , a continuous absorption process is known which decomposes the molecule into a normal and an excited atom. Usually no continuous absorption is known which corresponds to the dissociation of the molecule into normal atoms. For nitrogen we can expect the following process:

$$N_2(^1\Sigma_g) + h\nu = N^4S + N^3D - E_1,$$

$$N_2(^1\Sigma_g) + h\nu = N^4S + N^3P - E_2,$$

$$E_1 = D + N^3D - N^4S = 7.32 + 2.37 \text{ V} = 9.69 \text{ e-volts} = 1273 \text{ A},$$

$$E_2 = D + N^3P - N^4S = 7.32 + 3.66 \text{ V} = 10.98 \text{ e-volts} = 1124 \text{ A}.$$

No such continuous absorption corresponding to these processes has yet been discovered for N_2 . Hopfield noted two isolated absorptions at $\lambda 1518.02$ and $\lambda 1437.2$. The forbidden lines of nitrogen are

$$^4S-^3D = \nu \left. \begin{array}{l} 19223 \\ 19231 \end{array} \right\} \lambda \left. \begin{array}{l} 5202.1, \\ 5199.9, \end{array} \right.$$

$$^4S-^3P = \nu \quad \lambda \quad 3467.4,$$

$$^3D-^3P = \nu \left. \begin{array}{l} 9617 \\ 9606 \end{array} \right\} \lambda \left. \begin{array}{l} 10398.3, \\ 10410.2. \end{array} \right.$$

The last is in the infra-red and has not yet been observed either in the morning or the evening flash. $^4S-^3P$, $\lambda 3467.4$, is too far in the ultra-violet and has not yet been found. $^4S-^3D$, $\lambda 5202$, is promising, but it is unfortunately mixed up with the (0, 2) bands of N_2^+ . Sommer (Slipher and Sommer 1929) at one time identified the line observed near $\lambda 5206$ as being the $^4S-^3D$ line of N, but the matter needs careful reinvestigation. The discussion shows that as yet there is no definite evidence for the presence of forbidden lines of N in the night-sky spectrum, but this does not absolutely exclude the possibility of their occurrence. (Slipher (1933) has given some very strong red and infra-red bands, without indicating their origin.)

OXYGEN

Oxygen is very interesting, as so far no band lines due to O_2 or O_2^+ have been traced in the spectrum of the night sky or the polar aurora. The Fraunhofer spectrum of the sun shows only the A, B, ... bands of oxygen, and the night-sky spectrum shows only the green line ($^1D_2-^1S_0$), and the red lines due to the forbidden transition $^3P_{1,2}-^1D_2$. From these findings the conclusion has sometimes been drawn that oxygen exists in the upper atmosphere completely in the atomic state, i.e. is dissociated completely by sunlight above a certain height into atomic oxygen.

We get to a better understanding of these facts when we first critically consider the spectra of O_2 and O_2^+ , so far as they are known at the present time. Unfortunately, the knowledge, particularly beyond 1000 Å, is not as complete as can be desired. Whatever is known is presented in the energy-level diagram, fig. 2, and tables.

We observe that absorption by normal oxygen is confined to

- (1) The atmospheric bands corresponding to the transition

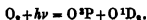
$$X^3\Sigma_g^- \leftarrow A^1\Sigma_u^-.$$

These bands occur very prominently in the Fraunhofer spectrum and constitute the A, B and α bands of Fraunhofer.

(2) The Schumann-Runge bands and the continuum beyond it. Of all the absorption bands of O_2 , these are the best studied. They correspond, according to Mulliken, to the transition

$$O(^3P + ^3P) X^3\Sigma_g^- \leftarrow O(^3P + ^1D_2) B^3\Sigma_u^-.$$

Beyond $\lambda 1750$, these bands pass into continuous absorption, corresponding, as Herzberg has shown, to the photochemical decomposition of O_2 into atoms as follows:



This absorption extends up to $\lambda 1250 \text{ \AA}$. The beginning at $\lambda 1750$ corresponds to $7.05 \text{ V} = D + E$, where D is the heat of dissociation of O_2 into two normal O^3P atoms, $= 5.09 \text{ V}$, and E is the heat of excitation of O^3P

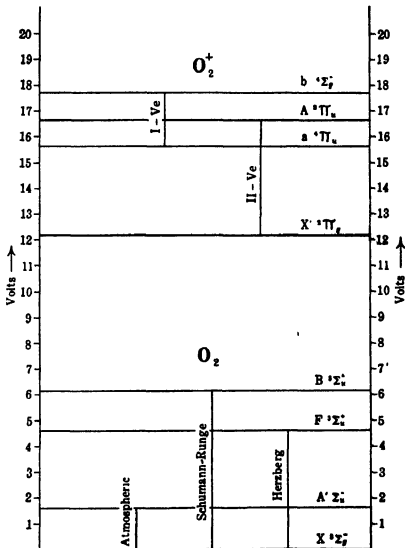


FIG. 2*

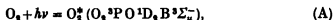
atom to the O^1D_2 state, viz. 1.96 V . For the lower wave-lengths in the continuum, absorption denotes that the products of decomposition separate with some kinetic energy.

* Reprinted from *Proc. Nat. Inst. Sci. Ind.* 1, 230, 1935.

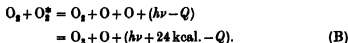
(3) A very feeble set of absorption bands, which appear to pass into a feeble continuum, has been reported by Herzberg at $\lambda\lambda 2595-2429$. The continuum at $\lambda 2429$ has been interpreted as giving rise to decomposition of O_3 into two normal O^3P atoms, for $\lambda 2429$ corresponds exactly to the dissociation potential of O_3 , viz. 5.09 V. But the intensity of this absorption is about 10^{-7} time less than that of the continuum beyond $\lambda 1750$, for a column of 0.002 cm. of O_3 at N.T.P. reduces $\lambda 1460$ to one-half, while 25 m. of O_3 at N.T.P. scarcely suffice to bring out the absorption of the bands at $\lambda\lambda 2595-2429$. According to Table I, this absorption may occur at about 20-40 km., but as this is the ozone region which cuts off all light between $\lambda\lambda 2900-2300$, we need not consider it.

The three types of absorption described above give rise to three well-known processes:

(a) The Schumann-Runge absorption ($X^3\Sigma_g^- \leftarrow B^3\Sigma_u^-$). This produces ozone as follows:



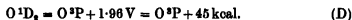
where O_3^* denotes the excited state of O_3 , viz. $B^3\Sigma_u^-$. The O_3^* molecule collides with a normal O_3 molecule (collision of the second kind) and dissociates it into two atoms as follows:



(b) The second process indicates that one of the atoms combines with O_3 to form an O_3 molecule. This process is highly probable, as the reaction is exothermic. We have, according to careful calorimetric investigations,



The energy set free is $h\nu + 24 - Q$, and if the wave-length used be $\lambda 2040$, $h\nu = 141 \text{ kcal.}$ Hence the energy set free is $141 + 24 - 117 = 48 \text{ kcal.}$, and this may raise the remaining O atom to the O^1D_2 state, as shown below:



This O^1D_2 atom reacts more readily with another O_3 atom and forms a second O_3 molecule and 72 kcal. of heat is produced. Thus, as Warburg experimentally established long ago, one quantum of light $< \lambda 2040$ produces two molecules of O_3 .

(c) Light of wave-length $> \lambda 2040$ can also convert O_3 into O_3 , but as has been shown experimentally, this is a high-pressure phenomenon. It corresponds to the Herzberg absorption at $\lambda 2425$. This probably plays no part in the formation of ozone in the upper atmosphere, and in the lower

atmosphere it is ineffective owing to the fact that $\lambda\lambda 3000-2300$ is cut off by ozone.

Let us now form a rough idea of the heights at which these processes take place in the upper atmosphere. *It is easy to show that the O_2 formation and the dissociation of O_2 into atoms* take place at quite different heights, and the two phenomena have no connexion with each other. According to Kreussler (see Lyman 1928), a length of 20 cm. of O_2 reduces $\lambda 1860$ (Schumann-Runge band) to two-thirds of its intensity. So according to Table I, $\lambda 1860$ is reduced to one-fourth of its intensity at a height of 45 km. But the other wave-lengths can be completely absorbed only at lower heights. $\lambda 1930$ loses 6.2 % of its intensity by the same column, so to reduce it to one-tenth its intensity, we require a column of oxygen 840 cm. in length, and this is reached at a height of about 35 km. These calculations are rather rough, but they suffice to show that the discovery of Götz, Meetham and Dobson (1934) that the O_2 layer exists at a height of 20-50 km. is in substantial agreement with the absorption data on Schumann-Runge bands. The decomposition of O_2 into O^3P and O^1D_2 takes place at quite different heights, for as mentioned before, $\lambda 1460$ is reduced to half its strength by 0.002 cm. of O_2 at N.T.P. So this ray is reduced to 10^{-8} time its strength when the equivalent column is 0.014 cm., and Table I shows that this is accomplished at about 180 km., i.e. above the E region of the ionosphere. The Herzberg absorption at $\lambda 2429$ appears to be ineffective in producing any substantial dissociation of O_2 into atoms, as this wave-length is absorbed by the ozone layer above a height of 30 km., above which there is not sufficient oxygen which can be appreciably dissociated in this fashion.

IONIZATION OF O_2 TO O_2^+

Let us now examine how O_2 is photo-ionized to O_2^+ . Tate and Smith (1932) have measured two ionization potentials of O_2 , viz. at 12.5 and 16.1 V. Mulliken and Stevens (1933) have shown how these measurements can be reconciled with the known spectroscopic levels of O_2^+ .

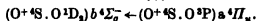
The known levels of O_2 are:

(1) The first negative band between $\lambda\lambda 5300-7900$ consisting of single headed bands shaded towards the red.

According to Mecke (1927) the band heads are given by

$$\nu = 16592.2 + (1180.3\nu' - 17.8\nu'^2) - (1026.1\nu'' - 11.1\nu''^2).$$

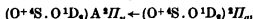
According to Mulliken and Stevens (1933) the electronic transition is



(2) The second negative bands between $\lambda\lambda 2200\text{--}4400$ consist of double bands shaded towards the red. The band heads are given by

$$\nu = 38308 + (887\nu' - 13.4\nu'^2) - (1859.9\nu'' - 16.53\nu''^2).$$

The electronic transition, according to Mulliken and Stevens, is



$r'' = 1.14, r' = 1.41$, hence bands like $(11, 0, 12, 0)$ lying between $\lambda\lambda 2100\text{--}1975$ are strong.

No intercombination between the two systems of bands is known.

The energy values of the levels are given by

Band	Ionization potential (experimental)	Spectroscopic
X ${}^2\Pi_g$	12.5	12.2
a ${}^4\Pi_u$	16.1	16.1
A ${}^2\Pi_u$	—	16.7
b ${}^4\Pi_g$	—	18.2

Recently a set of absorption experiments has been carried out by Price and Collins (1935) which show that the photoelectric ionization of O_2 to O_2^+ is effected in the same way as for N_2 , i.e. there is no ionization by light of wave-length corresponding to the lowest ionization potential of 12.5 V which can ionize O_2 to the lowest state of $O_2^+{}^2\Pi_g$, but they obtain a set of absorption bands (H, I, ...) which can be formed into a Rydberg sequence coming to a limit at $130,800\text{ cm.}^{-1}$, i.e. 16.14 V. This is identified with the a ${}^4\Pi_u$ state. Another set of bands passes to a continuum at 18.2 V. This is identified with ${}^4\Sigma_g$. A third strong ionization continuum is found at 16.7 V round about $\lambda 740$, which is provisionally identified with the production of the state A ${}^2\Pi_u$. There appears to be no absorption corresponding to the first observed ionization potential, viz. 12.5 e-volts. Thus it appears that the photochemically active light which can ionize O_2 to O_2^+ must have the energy equivalent to 16.20 e-volts at least, though the first ionization potential is 12.15 e-volts.

No negative bands of O_2^+ have yet been detected in the spectrum of the night sky or the aurora. Also, no photochemical process has so far been found which can be interpreted as decomposing the oxygen molecule into atoms, one of which is in the O^1S_0 state. In fact, our knowledge of the spectroscopy of O_2 and O_2^+ is meagre.

The value of the absorption coefficient at none of the supposed ionization continua has yet been obtained. It is stated by Price and Collins (1935) that a partial pressure of 0.01 mm. and a path of 1.5 m. (i.e. a path of 1.5

to 0.15 cm.) brings out the absorption quite clearly. According to Table I, we have this mass of oxygen at a height of about 100 km. It is therefore quite possible that the E layer ionization is produced by the ionization of O_2 to O_2^+ in the way supposed here.

The above summary of our knowledge of the spectra of O_2 and N_2 probably brings out clearly the necessity of undertaking a well-planned set of experiments on the spectroscopy of oxygen and nitrogen, particularly in the far ultra-violet, without which a satisfactory theory of the phenomena in the ionosphere, and of the luminous night-sky spectrum, is not possible.

Part of this work was done while the author was on tour in Europe and America as a Carnegie Fellow, and he wishes to acknowledge his indebtedness to the British Committee of the Carnegie Corporation for the award of the Carnegie grant to him. He also wishes to express his thanks to Professor E. A. Milne, F.R.S., and Professor H. H. Plaskett, F.R.S., for many useful discussions with them during his stay at Oxford. Thanks are also due to Professors H. Shapley and H. N. Russell, who were kind enough to discuss the contents of the paper with the author and offer many useful criticisms and suggestions.

SUMMARY

In this paper it is shown that a satisfactory theory of Upper Air Phenomena, such as the luminous night sky, the ionization in the different layers (E, F, ..), etc., must be based on a precise knowledge of the action of ultra-violet sunlight (below $\lambda 3000$) on molecular oxygen and nitrogen. The knowledge which we possess at present is summarized in the present paper; and the nature of molecular ionization and excitation, and the heights at which they occur, are discussed in detail. It is shown that according to laboratory evidence available at present molecular ionization of O_2 and N_2 by photochemical action does not take place at the lowest ionization potential, but at the second ionization potential, i.e. the photon which causes ionization of O_2 and N_2 leaves the molecular ion excited. It is concluded that the knowledge we possess is rather scanty and does not enable us to start any adequate physical theory, beyond indicating barely the physical processes which cause such phenomena. But the discussion brings out the necessity of carrying out well-planned laboratory experiments on the absorption spectra of O_2 and N_2 , without which it will not be possible to furnish an adequate explanation of upper air phenomena. The discussion also shows that the ultra-violet radiation from the sun differs widely from

that of a black body, and in selected wave-lengths the sun must be emitting nearly a million times more photons than is given by a black body at 6500° K. This may possibly be due to the fact that the ultra-violet spectrum of the sun may consist of a continuous background of faint light on which are superposed emission lines of H, He, He⁺, Fe⁺, and other elements which are represented in the visible range by lines of subordinate series, or by patches of ultra-violet continuous light (near about $\lambda 500$) leaking through the solar atmosphere from a much hotter region inside the photosphere, as suggested by Professor H. N. Russell.

REFERENCES

- Appleton, E. V. 1936 *Ann. Rep. Prog. Phys.* 2, 129-165.
Appleton, E. V. and Chapman, S. 1935 *Proc. Inst. Radio Engrs, N.Y.* 23, 658.
Brill 1932 *Handb. Astrophys.* 5 (1), 128-209.
Chapman, S. 1930 *Mem. R. Met. Soc.* 3, No. 26.
— 1931 *Proc. Roy. Soc. A*, 132, 353.
Dejardin 1936 *Rev. Mod. Phys.* 8, 1-30.
Fabry and Buisson 1913 *J. Phys. Radium*, 3, 196.
Fowler and Strutt 1917 *Proc. Roy. Soc. A*, 93, 577.
Götz, Meetham and Dobson 1934 *Proc. Roy. Soc. A*, 145, 416.
Gurney 1928 *Mon. Not. R. Astr. Soc.* 88, 377.
Hopfield 1928 *Phys. Rev.* 31, 1131.
— 1930 *Phys. Rev.* 36, 789.
Ladenburg, R. W. 1935 *J. Opt. Soc. Amer.* 25, 259.
Lyman 1928 "Ultra-violet Spectroscopy", p. 70.
Marrs and Hulburt 1929 *Phys. Rev.* 33, 412.
Mathur, L. S. and Sengupta, P. K. 1936 *Proc. Acad. Sci. U.P. Ind.* 5, 187-226.
Mecke 1927 *Z. Phys.* 42, 390.
Mitra and others 1936 *Proc. Nat. Inst. Sci. Ind.* 1, 131-59.
Mulders 1935 *Z. Astrophys.* 11, 132.
Mulliken 1933 *Phys. Rev.* 46, 144.
Mulliken and Stevens 1933 *Phys. Rev.* 44, 720.
Petit 1935 *Pub. Astr. Soc. Pacific*, 47, 324.
Price and Collins 1935 *Phys. Rev.* 48, 714.
Regener 1934 *Phys. Z.* 35, 788.
Saha 1935 *Proc. Nat. Inst. Sci.* 1, 238.
— 1936 *Science and Culture*, 1, 476.
Slipher 1933 *Mon. Not. R. Astr. Soc.* 93, 657.
Slipher and Sommer 1929 *Naturwissenschaften*, 17, 802.
Sponer 1933 "Molekülspektren", p. 17.
Störmer 1931 *Ergebn. koem. Phys.* 1, 1-82.
Tate and Smith 1932 *Phys. Rev.* 39, 270.
-

The Influence of Pressure on the Spontaneous Ignition of Inflammable Gas-Air Mixtures—The Simpler Olefines

BY G. P. KANE AND D. T. A. TOWNEND

(Communicated by W. A. Bone, F.R.S.—Received 10 December 1936)

INTRODUCTION

The investigations described in previous papers on this subject have related mainly to the paraffin hydrocarbons (Townend and Mandelkar 1933 *a, b*, Townend, Cohen and Mandelkar 1934; Townend and Chamberlain 1936, 1937). It has been found that mixtures with air of the members containing three or more carbon atoms, while not spontaneously ignitable at low pressures below about 500° C, give rise abruptly to ignition at higher pressures in a temperature range between about 310 and 370° C., where normally only cool flames are initiated; and although neither methane- nor ethane-air mixtures appear to develop cool flames, the latter are ultimately ignitable in a lower temperature system which is less complex than that characteristic of the higher paraffins.

Moreover, it is now recognized that "knock" in internal combustion engines arises in circumstances responsible for pronounced chemical reactivity in the unburnt explosive medium characteristic of that occurring in the lower temperature range (cf. Egerton and Ubbelohde 1935; Ubbelohde 1935), and the investigations referred to have indicated that the "knock-ratings" of the paraffins when used as fuels in such engines are related to the pressures requisite for the occurrence of spontaneous ignition in this range within an appropriate short time lag (Townend and Chamberlain 1936, p. 104, cf. Prettre 1936 *a* and *b*).

The purpose of this paper is to record our further observations with the simpler α -olefines: ethylene, propylene, butylene and amylene; meanwhile the nature of the complex chemical processes involved is forming the subject of separate investigations in these laboratories. The results of experiments in which small quantities of NO_2 and di-ethyl ether were added as promoters to representative propane- and propylene-air mixtures are also described.

The experimental procedure was essentially that described in previous papers.

RESULTS

A—Ethylene-Air Mixtures

The ethylene employed was supplied in cylinders by the British Oxygen Company and chemical analysis showed it to have a purity of 98.5 %, the residue apparently consisting of higher olefines. The influence of progressive increase of pressure on the ignition points was studied with four mixtures containing 3, 6, 10 and 20 % of the combustible (theoretical mixture for complete combustion = 6.5 %), respectively, and the results are illustrated by corresponding curves Nos. 1, 2, 3 and 4, fig. 1. No difficulty was ever encountered in reproducing the curves closely.

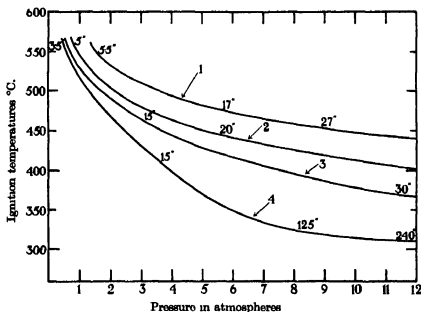


FIG. 1—Ethylene-air mixtures. Percentage mixtures for the curves: 1 = 3, 2 = 6, 3 = 10, 4 = 20.

The ignition points fell progressively with increase of pressure and smooth curves were obtained of the type previously observed with methane-air mixtures (Townend and Chamberlain 1936). The actual fall was, however, greater, the 20 % mixture being ignitable at 310° C. at a pressure of 11 atmospheres. There was never any indication of the formation of cool flames, nor of any irregularity either in the curves or in the pre-ignition time lags (small figures along the curves, fig. 1); the latter increased from about 4 sec. at 550° C. to about 240 sec. at 310° C., varying slightly with mixture composition.

It was not found possible to experiment with ethylene-oxygen mixtures undiluted with nitrogen owing to the extreme violence of the ignitions.

The fact that the ignition-point curves were quite smooth and that no superposed lower system as found with ethane-air mixtures (Townend and Chamberlain 1936) was observed is a matter of interest because strong evidence was adduced that with ethane this was attributable to an adequate formation and survival of acetaldehyde which is a principal intermediate product in its slow oxidation. If this be so, it seems possible that the known formation of acetaldehyde in the slow combustion of ethylene (Bone, Haffner and Rance 1933; Lehner 1932) arises not as a direct immediate oxidation product but as the outcome of a slower secondary process less likely to influence ignition. This would be in agreement with Bone's view (1933) that the primary product is vinyl alcohol, ethylene oxide and acetaldehyde arising by isomeric changes therefrom. Formaldehyde, on the other hand, which is known to promote the main combustion (Bone and others 1933), was always a product of the pre-flame combustion and the curves are consistent with its playing such a role.*

B—Propylene-Air Mixtures

The propylene used in these experiments was obtained from the Ohio Chemical and Manufacturing Company, Cleveland, U.S.A., its declared purity of 99.5 % being confirmed by chemical analysis. Determinations of the ignition points were restricted to three mixtures containing 3, 4.5 and 6 % (theoretical mixture = 4.43 %) of the combustible, curves Nos. 1, 2 and 3, fig. 2A, respectively.

The influence of pressure was found to be much more marked than with ethylene-air mixtures. The ignition points of all three mixtures fell progressively at the lower pressures to about 440° C.; further increase of pressure, however, caused an abrupt fall over a temperature range where cool flames (shaded areas) were also observable, but below about 330° C. increase in pressure had little further influence. The time lags were somewhat longer with propylene- than with ethylene-air mixtures at the higher temperatures but were of much the same magnitude in the lower range. Moreover, although the ignition points fell rapidly no irregularities in the time lags were observed, in this respect and in the fact that the curves showed no pressure minima in the cool-flame range, propylene differs from propane (Townend and Chamberlain 1936). The main difference in behaviour, however, lay in the weakness both of the pressure pulses and the luminosity

* Cf. the influence of formaldehyde addition to methane-air mixtures (Townend and Chamberlain 1936).

associated with the cool flames, a fact which made it impossible to determine pressure and temperature limits for them with precision. A further observation was that although for corresponding mixtures those with propylene were ignitable in the lower system at slightly lower pressures than those of propane, not only were the time lags longer with the former but their rate of decrease with increase of pressure was not as great; curves Nos. 4 and 5, fig. 2, relate to ignitions of the 4.5 % mixture No. 2, with a constant lag of 3 and 2 sec., respectively, and may be compared with the curves previously published (Townend and Chamberlain 1936) for a 7.5 % propane-air mixture.

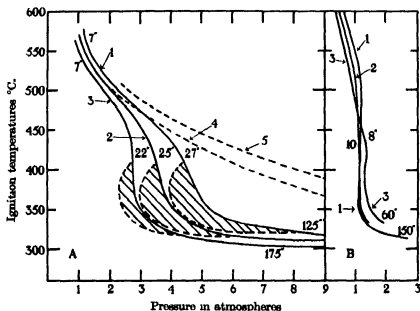


FIG. 2—A, propylene-air mixtures. Percentage mixtures for the curves: 1 = 3.0, 2 = 4.5, 3 = 6. Curves 4 and 5, for the 4.5 % mixture, relate to ignition with constant lags of 3 and 2 sec. Shaded areas indicate cool-flame regions. B, propylene-oxygen mixtures. Percentage mixtures for the curves: 1 = 25, 2 = 18, 3 = 10.

So far in these investigations it has not been possible to contribute greatly to any kinetic study of the problem because of the violence usually associated with the ignition of undiluted mixtures; hence hydrocarbon-air media have mainly been studied in the first place, because of their direct relation with the problem of "knock". An attempt was made to study the behaviour of propylene-oxygen mixtures, however, for the time lags were long enough to allow of the mixture being safely admitted to the explosion vessel and the valves then closed before ignition occurred. It was found impossible to employ mixtures with more than 25 % of combustible because of copious

carbon deposition which rendered the ignitions unreproducible; three mixtures of composition $C_3H_8 + 3O_2$; $C_3H_8 + 4.5O_2$ (theoretical mixture) and $C_3H_8 + 9O_2$ were used successfully.

The general slope of the curves (fig. 2 B) simulated that observed with the propylene-air mixtures, an interesting feature, however, is that they intersect, the concentration of hydrocarbon having a predominating influence below $420^\circ C$, above which the reverse was the case, indicating with these mixtures the intervention of other influences in the higher temperature range. Moreover, the ignitions were much more violent in the lower range and akin to detonation, the time lags were also relatively shorter than for corresponding mixtures with air and increased progressively from a fraction of a second at $600^\circ C$. to $2\frac{1}{2}$ min. at $316^\circ C$.

C—Butylene-Air Mixtures

Butylene is the lowest olefine to exhibit isomerism; in order to obtain a comparison in behaviour with the normal paraffins it was essential to experiment with α -butylene ($CH_3:CH.CH_2.CH_3$) which has the longest saturated chain. Difficulties were experienced in achieving this end, as the isomers have similar properties and isomerism occurs if high working temperatures are employed in synthesizing the material. The physical properties of the three butylenes as given by Coffin and Maaß (1928) suggested a method for separating α -butylene, it being the only hydrocarbon likely to be met with in any mixture still liquid at the temperature of liquid air.

The best source of supply of the olefine was found to be a crude butylene supplied by the Ohio Chemical and Manufacturing Company and containing 85 % of a mixture of the three butylenes together with butane, propane and propylene. An apparatus was devised whereby the material was first liquefied at $-78^\circ C$. and the more volatile impurities (e.g. propane and propylene) removed by brief evacuation, and the residual liquid subsequently allowed to boil over into a vessel maintained at liquid-air temperature. While the remaining impurities solidified in this vessel the α -butylene remained liquid and was sucked into a further vessel also at liquid-air temperature. The process was repeated many times in order to obtain an adequate supply of the olefine; the accumulated product was refractionated and collected in a cylinder. It condensed finally in liquid air to a clear viscous liquid, and determinations both of its vapour pressure and the m.p. of its dibromo derivative showed that it was a good specimen of α -butylene.

Owing to the shortage of combustible the experiments were restricted

to two mixtures containing 3.36 % (the theoretical mixture) and 5.5 % of it, respectively. The ignition-point curves determined in the usual manner are reproduced in fig. 3 A, Nos. 1 and 2. They resemble closely those

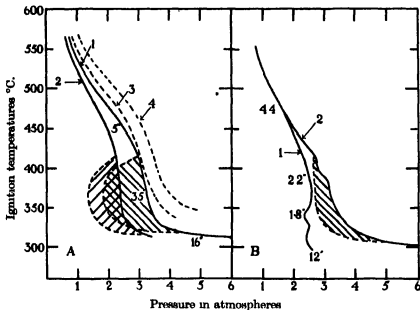


FIG. 3—A, butylene-air mixtures. Percentage mixtures for the curves: 1 = 3.36, 2 = 5.5. Curves 3 and 4, for the 3.36 % mixtures, relate to ignition with constant lags of 3 and 2 sec. Shaded areas indicate cool-flame regions. B, amylene-air mixtures. 1 = 2 % α -amylene, 2 = 2 % isomeric amylenes.

obtained with propylene-air mixtures, the ignition points being somewhat lower at the same pressures and the mixtures ignitable at lower pressures in the lower temperature range. Although no indication of the pressure minima of ignition in the lower temperature range characteristic of the higher paraffins was observed, the lags, which were much shorter than those observed with either ethylene- or propylene-air mixtures, showed a comparable irregularity as found with the paraffins; thus they were reduced from about 5 sec. at 450° C. to a minimum value of 3.5 sec. at 370° C., thereafter increasing again to 16 sec. at still lower temperatures, as the minimum pressures were further increased. The cool flames were also more pronounced both in regard to the accompanying pressure pulses and their luminosity.

D—Amylene-Air Mixtures

Amylene as ordinarily supplied commercially has a b.p. range of from 36 to 40° C. According to Wilkinson (1931) the b.p. of α -amylene is 30.3° C.

It was decided to synthesize the olefine by means of his method, which employs the Grignard reaction:



The yield of pure material was low and barely sufficient was available to complete the ignition-point curve of one mixture containing 2% of it (theoretical mixture = 2.7%), No. 1, fig. 3B. For our information other curves were determined with a sample of purchased amylene containing mainly more stable isomers, and curve No. 2 relates to a 2% mixture of this material.

The ignition points of the α -amylene mixture (No. 1) gave for the first time definite evidence of the pressure minima in the cool-flame temperature range, invariably found with the higher paraffins. The higher minima occurred at 340° C. and the lower at about 312° C., but lack of material prevented close repetition of this part of the curve. The time lags were also irregular as with the higher paraffins, a minimum value of 1½ sec. being observed at 340–350° C. Cool flames were observable between 310 and 410° C.; the pressure pulses being marked between 310 and 360° C.

As might be anticipated, the mixture of the side-chained isomers (b.p. 36–40° C.) with air (No. 2, fig. 3B) was not nearly as easily ignitable as the α -olefine; this stresses the importance when comparing the behaviour of members of one series with those of another, of employing corresponding isomers, a matter which is sometimes overlooked.

Summary—Summarizing the results observed with the four simpler olefines and their relation with the problem of "knock", it may be stated that the influence of pressure on the spontaneous ignition of their mixtures with air is in the main not unlike that previously observed with the paraffins, as can be seen from the following:

(1) While the ignition points of ethylene-air mixtures fell smoothly with increase of pressure, as was the case with methane-air mixtures, and did not develop a lower ignition system, the curves for the mixtures of the higher olefines exhibit a sharp fall over narrow critical pressure ranges to a lower system where cool flames are normally observed.

(2) Although the curves for propylene- and butylene-air mixtures did not reveal the existence of negative temperature coefficients as manifest with the corresponding paraffins by the existence of pressure minima of spontaneous ignition in the lower temperature system, that for a 2% α -amylene-air mixture showed characteristics typical of the higher paraffin-air mixtures, in that two pressure minima were located at corresponding temperatures.

(3) The minimum pressures requisite for ignition in the low temperature range with corresponding mixtures, considered also in the light of the observed time lags, indicate that the "knock-ratings" of the olefines should fall as the series is ascended. This is in accord with the "Critical Compression Ratios" of Boyd and his collaborators (Spiers 1935), as the figures in Table I show.*†

TABLE I

Hydrocarbon	Critical Compression Ratio	Minimum pressure at 350° C. (atm.)	Time lag (sec.)
Ethylene	8.5	—	—
Propylene	8.4	5.0	30
α -Butylene	—	3.35	3.5
α -Amylene	5.8	2.4	1.8

Indeed, our results as a whole are in conformity with the general conclusions of Boyd, that (1) in an homologous series, whether paraffinic or olefinic, the tendency to "knock" increases with the increasing length of the saturated carbon chain, and (2) in an isomeric series, "knock" decreases progressively with the centralization of the double bond.

With regard to the complex chemical processes involved in the combustion of either paraffins or olefines containing more than three carbon atoms, at least three conditions have to be considered, namely: (1) that giving rise to cool flames somewhere in the temperature range between about 280 and 410° C.; (2) that ultimately resulting in true ignition in the products after the passage of cool flames; there being usually two distinct zones of activity between (a) 300–330° C., and (b) 340–370° C., respectively; and (3) that occurring above the upper temperature limit for cool flames, i.e. above about 410° C. There is now little doubt that conditions conducive to (1) are mainly responsible for "knock" and that the phenomenon itself

* It does not correspond, however, with the "blending octane numbers" of Garner and co-workers (1933) which indicate that α -butylene should exhibit a maximum "knock-rating".

† The values for corresponding mixtures of the paraffins, now brought to date, with which those in Table I may be compared are as follows:

Hydrocarbon	C.C.R.	Minimum pressure (atm.)	Time lag (sec.)
Ethane	14	—	—
Propane	12	6.8	3
<i>n</i> -Butane	6.4	2.7	2
<i>n</i> -Pentane	3.8	2.2	1.4
<i>n</i> -Hexane	3.3	1.9	1

involves essentially the processes occurring in (2) which become very intense at sufficiently high pressures.

There is also little doubt that the processes occurring in cool flames are of a chain character and probably associated with the presence of peroxides (cf. Aivazov and Neumann 1936*a*), aldehydes or other active intermediate bodies. With the simpler hydrocarbons, aldehydes are important intermediate products in this respect (Townend and Chamberlain 1936), but with higher members other bodies are probably concerned. Egerton and his collaborators (1935) have considered primary peroxides as mainly concerned and the recent discovery of Newitt (1936) that mixed ethers are formed in the slow combustion of olefines has revealed another class of intermediate compounds, the possible influence of which has called for investigation.

No complete explanation has so far been forthcoming to account for the upper temperature limit for cool flames. Opinion generally favours the view that some material whose survival is vital to the low-temperature combustion becomes thermally unstable; other explanations are equally attractive, however, and much further work is necessary.

While the present investigation has shown the behaviour of the α -olefines under pressure to be not unlike that of corresponding paraffins, the main difference lies in the reduced luminosity of the cool flames and the feeble pressure pulse connected with them, indicating that the processes operative in the low-temperature system are less intense than with the paraffins; moreover, the pre-flame time lags are not only greater but decrease less rapidly with pressure, although with propylene the minimum ignition pressures are less than those for propane.

Recently Semenoff (1935*a*) has accounted for spontaneous ignition as arising in certain circumstances by a process of interacting chains, which becomes of importance when the number of active centres (usually developed by degenerate branching) reaches an adequate concentration, in accordance with the expression

$$\frac{dn}{dt} = n_0 r - (g-f)n + f'n^2,$$

where n_0 = the number of active centres formed, r = the length of the primary chain, g = the chain-breaking factor, f = the chain-branching factor, and $f'n^2$ takes account of chain interaction; f' being small.

Aivazov and Neumann (1936*b*) have also developed an empirical equation based on this view for the relation between the pre-ignition time lags for cool flames, which are regarded as ignitions giving rise to stable intermediate

products, and the experimental temperature, pressure, vessel diameter and minimum cool-flame pressure. This expression has been found to yield values in close agreement with their own experimentally determined lags for pentane-oxygen mixtures and those previously published from these laboratories for propane-air mixtures.

The degenerate branching view of chain development has some attraction when comparing the respective behaviour of cool-flame formation with corresponding simpler paraffins and olefines; for with the former the lags are much shorter and the cool flames more intense (cf. Semenoff 1935*b*). This may be connected not so much with any difference in the character of the chain initiators (Ubbelohde 1935, p. 371) in the two combustions as with possibly either a slower building up of active centres by virtue of difference in the chemical stages involved, or the intervention of stronger inhibiting processes in the case of the olefines.

E—Experiments with Media Sensitized by NO₂ and Di-ethyl Ether

Hitherto with higher paraffin-air media we have studied the promoting influence only of higher aldehydes (particularly acetaldehyde) likely to be formed during their slow combustion, finding generally that while reaction in the low temperature system is enhanced, at higher temperatures the effect diminishes, becoming negligible above about 500° C. (cf. Aivazov and Neumann 1936 *a* and *b*). That this influence is general may be seen from curve 4, fig. 4B, which shows the influence of the addition of 1% acetaldehyde to a 4.5% propylene-air mixture (curve 1).

NO₂ additions—Recently we have studied the influence of small additions to the explosive media of NO₂ and ether. In fig. 4A the influence of successive additions of 0.1 and 1% of NO₂ to a 7.5% propane-air mixture (curve No. 1) may be seen from the corresponding ignition-point curves (Nos. 2 and 3 respectively). In the main it was found that by far the greatest influence of the presence of NO₂ was not only to reduce the time lags, but to effect ignition at much lower pressures in the temperature range above the upper limit for cool flames. In fact there now appeared to be an abrupt transition from ignition arising from the sensitized chemical processes typical of the upper system to ignition arising as a sequence of cool-flame propagation, as is indicated by the dotted line (curve No. 3).

In the low-temperature range, however, the presence of NO₂ while reducing the time lags, had by comparison no great influence in extending either the temperature or pressure limits within which cool flames are propagated; moreover, the minimum pressures at which true ignition was

superposed on cool flames were only slightly reduced as may be seen from the curves.

A likely interpretation of these results would seem that the cool-flame processes arise in circumstances responsible for the preservation of some intermedially formed chain initiating species, which is equally as potent a promoter of the combustion as NO_2 . The higher ignition system, being conditioned by circumstances unaided by the survival of this material, is

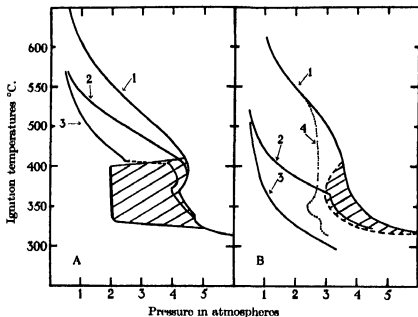


FIG. 4.—A, curve 1 = 7.5% propane-air mixture, 2 and 3 are the same but with 0.1 and 1.0% additions of NO_2 . B, curve 1 = 4.5% propylene-air mixture, 2 and 3 are the same but with 0.1 and 1.0% additions of NO_2 , and 4 with 1.0% CH_3CHO .

much more sensitive to the chain initiating propensity of NO_2 . It is also conceivable that the presence of NO_2 may lead to some mutual destruction of reactive centres in the lower system. This comparative ineffectiveness of NO_2 in promoting combustion in the cool-flame range is also of great interest because it is in keeping with the observation of Egerton, Ll. Smith and Ubbelohde (1935) that it had no influence in inducing "knock" with butane, heptane, and petrol, although with hydrogen it did so.

In fig. 4B, a corresponding set of curves has been drawn showing the respective influence of the addition of 0.1 and 1% of NO_2 (curves 2 and 3) to a 4.5% propylene-air mixture (curve 1). Although a much greater reduction in the relevant time lags was found, the presence of NO_2 in the medium was without any appreciable influence on the pressure and temperature

limits within which cool flames were observed. With regard to true ignition however, the presence of 0.1 % of NO_2 not only very greatly reduced the minimum ignition pressure in the upper system, but promoted ignition resulting in the lower system in the cool-flame products (curve 2); and 1 % of it so sensitized the combustion that a smooth curve was obtained, ignition occurring over the whole range without any apparent prior cool-flame reaction whatever.

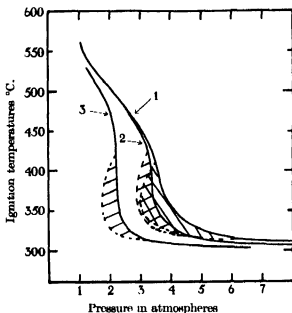


FIG. 5—Curve 1 = 4.5 % propylene-air mixture, 2 and 3 are the same but with 0.1 and 0.5 % additions of $(\text{C}_2\text{H}_5)_2\text{O}$. Shaded areas indicate cool-flame regions.

The marked difference in the relative influence of NO_2 in the lower system with propane and with propylene seems in general agreement with the expression derived for the influence of the initial concentration of active centres on the rate of reaction by Aivazov and Neumann (1936 *b*), for the rate of chain development with propylene, which is by comparison with propane a slow process, would be the more enhanced by the addition of active centres to the system.

Ether Additions—As far as we are aware, the influence of ether as a promoter of combustion has not been previously studied, although Egerton and his collaborators found diethyl peroxide to be an active "pro-knock". In fig. 5 curves have been drawn showing the influence of successive additions of 0.1 and 0.5 % of it (curves 2 and 3) to a 4.5 % propylene-air mixture.

Speaking generally ether was found to play a role rather like, but more powerful than, that of acetaldehyde (see Townend and Chamberlain 1936), for 0.1 % of it, while facilitating both the initiation of cool flames and true ignition in the lower temperature range below 425° C., was without influence at higher temperatures (curve 2); an addition of 0.5 %, however, powerfully promoted the combustion (curve 3), this being marked even up to 500° C. So far as the matter has been tested, a like effect was found with a 7.5 % propane-air mixture.

The influence of ether as a promoter is thus very striking, but it would be premature at this juncture to form any definite opinion as to the precise role played by it, since the combustion of ether itself has been found to be a very complex process (Townend and Chamberlain 1937).

In conclusion, we desire to express our thanks to the University of Bombay, India, for a Sir Mangaldas Nathubhai Technical Scholarship, which has enabled one of us (G. P. K.) to devote his whole time to the work.

SUMMARY

Previous investigations into the influence of pressure on the spontaneous ignition of the paraffin hydrocarbons have been extended to the simpler olefines, ethylene, propylene, α -butylene and α -amylene. The ignition-point curves for members with more than three carbon atoms are not unlike those of corresponding paraffins, for a sharp fall was again observed over narrow critical pressure ranges to a lower temperature system, where cool flames are normally observed; and with a 2 % amylene-air mixture two pressure minima in the lower system were located as with the higher paraffins. The main difference between the two series lies in the longer pre-flame time lags observed with the olefine-air media and in the weak luminosity and feeble pressure pulses associated with the cool flames to which they give rise.

Further experiments are described in which small amounts of NO_2 and di-ethyl ether were employed as promoters.

REFERENCES

- Aivazov and Neumann 1936a *Acta phys. chim. U.S.S.R.* 4, 575.
— — 1936b *Z. phys. Chem.* B, 33, 349.
Bone 1933 *J. Chem. Soc.* p. 1599.
Bone, Haffner and Rance 1933 *Proc. Roy. Soc. A*, 143, 16.
Coffin and Määs 1928 *J. Amer. Chem. Soc.* 50, 1427.

- Egerton, Ll. Smith and Ubbelohde 1935 *Philos. Trans. A*, **234**, 433.
 Garner and others 1933 *Proc. World Petr. Conf.* **2**, 170.
 Lehner 1932 *J. Amer. Chem. Soc.* **54**, 1830.
 Newitt 1936 *Proc. Roy. Soc. A*, **157**, 348.
 Prettre 1936 *a C.R. Acad. Sci., Paris*, **202**, 1178.
 — 1936 *b Ann. Combustibles Liquides*, **11**, 669.
 Semenov 1935 *a Z. phys. Chem. B*, **28**, 43.
 — 1935 *b "Chemical Kinetics and Chain Reactions,"* p. 73.
 Spiers 1935 "Technical Data on Fuel."
 Townend and Chamberlain 1936 *Proc. Roy. Soc. A*, **154**, 95.
 — — 1937 *Proc. Roy. Soc. A*, **158**, 415.
 Townend, Cohen and Mandelkar 1934 *Proc. Roy. Soc. A*, **146**, 113.
 Townend and Mandelkar 1933 *a Proc. Roy. Soc. A*, **141**, 484.
 — — 1933 *b Proc. Roy. Soc. A*, **143**, 168.
 Ubbelohde 1935 *Proc. Roy. Soc. A*, **152**, 354.
 Wilkinson 1931 *J. Chem. Soc.* p. 3057.

A Criticism of the Method of Expansion in Powers of the Gravitational Constant in General Relativity

BY J. L. SYNGE

(Communicated by E. T. Whittaker, F.R.S.—Received 14 December 1936)

In a recent paper by Temple (1936) there is employed a method based on expansions in powers of the gravitational constant. It is the purpose of the present note to show that the method is fundamentally unsound.

Temple considers, not a single universe, but a singly infinite set of universes in which the gravitational constant γ takes on a range of values. Let us denote a member of this set by $V_4(\gamma)$. Let there be a set of observers $S(\gamma)$, one for each universe, for we cannot speak of a single observer observing two different universes. Each of these observers $S(\gamma)$ assigns a system of co-ordinates x^i in his particular universe $V_4(\gamma)$; thus the set of observers $S(\gamma)$ assigns co-ordinates x^i to all the events in the set of universes $V_4(\gamma)$. Then, since the gravitational constant is involved in the field equations, the components g_{ij} of the fundamental tensor will be functions of x^i and γ . Let them be expanded in power series

$$g_{ij} = \alpha_{ij} + \gamma\beta_{ij} + \dots, \quad (1)$$

where α_{ij} , β_{ij} , ... are functions of the co-ordinates only.

Consider now a completely independent set of observers $S'(\gamma)$ in the same set of universes $V_4(\gamma)$. The set of observers $S'(\gamma)$ assigns co-ordinates x'^i to all the events in the set of universes $V_4(\gamma)$, and the components of the fundamental tensor g'_{ij} are expanded in power series

$$g'_{ij} = \alpha'_{ij} + \gamma \beta'_{ij} + \dots \quad (2)$$

The transformation between the co-ordinates assigned by the set of observers $S(\gamma)$ and those assigned by the set of observers $S'(\gamma)$ will be of the form

$$x^i = x^i(x'^1, x'^2, x'^3, x'^4, \gamma), \quad (3)$$

and *not* in general of the form

$$x^i = x^i(x'^1, x'^2, x'^3, x'^4). \quad (4)$$

For suppose that the transformation is of the form (4). Let P_1, P_2 be events in the universes $V_4(\gamma_1), V_4(\gamma_2)$, respectively, to which the observers $S'(\gamma_1), S'(\gamma_2)$ have assigned the same co-ordinates x'^i . Then, by (4), it follows that the observers $S(\gamma_1), S(\gamma_2)$ assign the same co-ordinates x^i to P_1, P_2 . But $S(\gamma_1), S(\gamma_2), S'(\gamma_1), S'(\gamma_2)$ are four observers operating with complete independence. It will not happen in general that $S(\gamma_1), S(\gamma_2)$ assign the same co-ordinates x^i to two events P_1, P_2 , picked at random from their respective universes. Hence the transformation is in general of the form (3), not (4).

Now
$$g'_{ij} = g_{kl} \frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j}, \quad (5)$$

for all values of γ ; hence, letting γ tend to zero,

$$\alpha'_{ij} = \alpha_{kl} \left(\frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} \right)_{\gamma=0}. \quad (6)$$

Thus α'_{ij} are the components of a tensor in the universe $V_4(0)$, but not in the universes $V_4(\gamma)$ for $\gamma > 0$.

Differentiating (5) with respect to γ , and letting γ tend to zero, we get

$$\beta'_{ij} = \beta_{kl} \left(\frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} \right)_{\gamma=0} + \alpha_{kl} \left[\frac{\partial}{\partial \gamma} \left(\frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} \right) \right]_{\gamma=0}. \quad (7)$$

Thus β'_{ij} are not the components of a tensor in any of the universes $V_4(\gamma)$ ($\gamma \geq 0$).

The above argument applies of course to the expansions of other tensors as well as g_{ij} .

Consequently Temple's form of Gauss's theorem is not valid, for it rests on an assumption as to tensor character which is not valid. The intrinsic teleparallelism described in his paper does not exist.

That does not mean that there is no intrinsic teleparallelism in Riemannian space. Temple's statement that "... the absence of any teleparallelism is the most characteristic feature of abstract Riemann space" is incorrect, if by teleparallelism we understand the existence in a space V_N of N orthogonal unit vectors intrinsically defined at each point. If S_{ij} is any symmetric covariant tensor of the second order, intrinsically defined, the equations

$$S_{ij}\lambda^j + \theta g_{ij}\lambda^j = 0, \quad (8)$$

where θ satisfies the determinantal equation

$$|S_{ij} + \theta g_{ij}| = 0, \quad (9)$$

define a set of N orthogonal vectors λ^i , provided that (9) has N distinct real roots. This process was given by Ricci (1904; cf. Eisenhart 1926, p. 114) with $S_{ij} = R_{ij}$, the Ricci tensor. In space time unoccupied by matter ($T_{ij} = 0$) these principal directions become indeterminate, but there are other principal directions which remain determinate unless the space has further special symmetry (Synge 1922).

SUMMARY

In a recent paper Temple has employed a method based on the expansion of tensors in space time in powers of the gravitational constant. The assumption that the coefficients in such power series are themselves tensors forms the basis of his definition of teleparallelism and of his form of Gauss's theorem. It is shown in the present note that the coefficients in question do not possess tensor character, a fact which invalidates Temple's results. The absence of tensor character is due to the fact that there is under consideration, not one universe, but a set of universes, and hence we must take into consideration transformations of co-ordinates which involve the gravitational constant.

REFERENCES

- Eisenhart 1926 "Riemannian Geometry." Princeton.
 Ricci 1904 *Atti Ist. veneto*, 63, 1233.
 Synge 1922 *Proc. nat. Acad. Sci.*, Wash., 8, 198.
 Temple 1936 *Proc. Roy. Soc. A*, 154, 354.

The Scattering of Alpha Particles in Helium, Hydrogen and Deuterium

BY C. B. O. MOHR, Ph.D., *Exhibition of 1851 Senior Student,*
Trinity College, Cambridge

AND G. E. PRINGLE, B.A., *Emmanuel College, Cambridge*

(Communicated by Lord Rutherford, O.M., F.R.S.—
Received 17 December 1936)

1—INTRODUCTION

The study of the scattering of α -particles is now a familiar method of finding the fields of force of light nuclei, because the inverse square law of repulsion breaks down at distances to which α -particles are able to approach them. It has been shown how wave mechanics may be applied to the interpretation of experimental data in terms of an interaction potential between the nuclei which deviates from the Coulomb value at separations of the order 3×10^{-13} cm. This study is valuable, for with knowledge about the simplest particles one may more easily approach the difficult problem of the structure and stability of the heavier nuclei.

The previous experimental work on the scattering of α -particles in helium by Rutherford and Chadwick (1927), and by Chadwick (1930), and in hydrogen by Chadwick and Bieler (1921) cannot be regarded as sufficiently precise, and for deuterium we have only the preliminary experiments of Rutherford and Kempton (1934), and of Pollard and Margenau (1935). Wright's (1932) results for helium are more definite, but even they do not complete the desired information for the slower α -particles. For hydrogen and deuterium no means were used to obtain accurate absolute values.

A theoretical method for deducing from a set of experimental results the form of the nuclear potential at close distances of approach was first given by Taylor (1931, 1932). The potential derived by him and given in his paper is probably not very reliable however, because it was based on rather unreliable scattering data. His result must also be held suspect because it leads to a large binding energy for ${}^6\text{Be}$ which can hardly be correct. Further, in view of the results of this paper it is possible that his method cannot be applied so safely as was then thought. The method requires the de Broglie wave-length of the relative motion to be large compared with nuclear dimensions, that is with the greatest distance at which significant deviations

from the Coulomb law are found. This assumption is more likely to be justified for He-H collisions than for He-He, and may have been at fault for the latter.

We therefore considered it necessary to do systematic experiments with α -particles of polonium on the three lightest nuclei, helium, hydrogen and deuterium, with special attention to the angular distribution, using modern counting methods, and interpreting the results in terms of a simple form of nuclear potential of the Gamow type. The most interesting feature which appears from this work is that the radii* of the lightest nuclei are over 10^{-13} cm., a value which is some three times larger than was previously supposed, and agrees better with early estimates made on the classical theory.

2—EXPERIMENTAL PROCEDURE

The experiments resembled Wright's except in detail, and the procedure is described in his paper. Only small modifications were required in using hydrogen and deuterium instead of helium. When the former two gases were used, the recoil protons and deuterons projected by the α -particles were counted. It is not a satisfactory method to count the scattered α -particles, for these are confined to small angles with the incident beam, viz. less than $14\frac{1}{2}^\circ$ for hydrogen and 30° for deuterium. If one counts the recoil

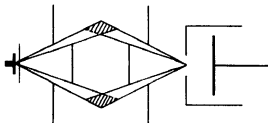


FIG. 1

nuclei it is, in fact, much easier to interpret the results, and one can readily distinguish them from the larger impulses produced by those few α -particles which are scattered into the counter.

The apparatus included a scattering chamber constructed for the "angular ring" method of observing α -particles scattered in gases (see fig. 1). The carbon diaphragms and disks defining the incident and scattered beams

* We shall use the term "radius" to denote the distance between two nuclei at which appreciable departure of the field from its Coulomb value first takes place.

were mounted in such a way that they could be quickly but accurately aligned with the source and counter. The geometrical arrangement was slightly better than that of Wright's apparatus.

In order to determine the ratio of the actual to the "classical" scattering, argon was substituted in the scattering chamber, and the scattered α -particles were counted. When the pressures were not exactly "equivalent" as regards stopping power, a correction was applied. The accuracy of the correction was carefully tested by experiments with different pressures of argon in the scattering chamber.

A small ionization chamber was used as a linear counter. When helium was used, no window was placed over the hole in the front of the counter, and the gas itself was present at sufficient pressure to make counting possible. When argon was substituted, no difficulty was experienced in counting with comparatively low pressures. With this arrangement particles of short range were retarded as little as possible before reaching the counting space, and those entering obliquely were also counted. Wright, in his experiments, used a ball counter, but this has neither of the above advantages. In order to count the weaker impulses of the recoil particles more easily, a window was placed over the hole in the front of the counter which was then filled with argon at atmospheric pressure.

The difficulty caused by the production of hydrides when polonium is exposed to hydrogen was avoided by admitting air to an isolated system which included the source. The gases were admitted through leaks to prevent pressure differences which would have broken the mica window in front of the source. In all cases, the path from source to chamber was as long as possible and there was no difficulty due to progressive contamination. The helium was purified before each experiment by passing it through charcoal cooled in liquid air, and care was taken to see that there was no air leaking into the apparatus. The amount of impurity in the hydrogen and the deuterium was less than 1 %.

The amplifier was of the type described by Wynn-Williams and Ward (1931) with a pentode valve in the last stage, operating either a moving-iron oscillograph, or headphones when this was more convenient. The number of particles counted for each point was generally sufficient to bring the probable error below 10 %.

The range of the incident α -particles was varied by placing in front of the source thin sheets of mica carried on disks which could be rotated from outside the scattering chamber. In most of the experiments the sources were of 2-4 millicuries of polonium; for preparing them we are indebted to Professor Chadwick. The sources were never long exposed to the air while the

experiments were in progress, so that tarnishing of the metal disk was minimized.

3—RESULTS

The ratio, R , of the intensity of scattering to the classical value was determined for angles of scattering of 15° , 21° , 27° , 33° , 39° , 45° in helium, and for recoil angles of 15° , 21° , 33° , 45° in hydrogen and in deuterium. The range of angles, as measured from the centre of the source to the centre of the counting-area, was as follows: $15^\circ \pm 2^\circ$, $21^\circ \pm 4^\circ$, $27^\circ \pm 4^\circ$, $33^\circ \pm 5^\circ$, $39^\circ \pm 7^\circ$, $45^\circ \pm 4^\circ$. The results are given in Tables I, II and III, and plotted in figs. 2, 3 and 4. The main feature to be emphasized at once is that Coulomb scattering is not observed until the range of the α -particles is made much less than has hitherto been thought necessary.

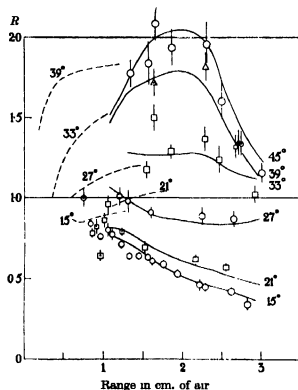
TABLE I—THE RATIO OF OBSERVED TO CLASSICAL SCATTERING IN HELIUM

ϕ	Range mm.	" R "	P. E.	ϕ	Range mm.	" R "	P. E.
15°	7.5	1.00	0.06	27°	12.1	1.01	0.06
	8.4	0.84	0.03		13.2	0.98	0.08
	9.6	0.76	0.02		16.1	0.91	0.04
	10.5	0.80	0.05		22.5	0.89	0.06
	11.1	0.77	0.04		26.5	0.87	0.05
	12.2	0.71	0.03	33°	15.5	1.18	0.07
	13.3	0.64	0.02		16.4	1.50	0.09
	14.5	0.64	0.02		18.6	1.29	0.05
	15.6	0.63	0.02		22.9	1.37	0.08
	16.2	0.61	0.03		24.7	1.24	0.09
	17.6	0.59	0.03		26.8	1.32	0.08
	19.4	0.53	0.02		29.2	1.02	0.06
	22.2	0.46	0.04	39°	13.7	1.33	0.07
	22.9	0.45	0.02		16.5	1.72	0.09
	26.2	0.42	0.03		23.1	1.82	0.09
	28.3	0.34	0.04		27.2	1.35	0.08
21°	8.6	0.78	0.03	45°	13.5	1.78	0.09
	9.1	0.82	0.05		15.8	1.84	0.14
	9.6	0.64	0.04		16.7	2.09	0.11
	10.1	0.86	0.05		18.8	1.94	0.12
	10.6	0.96	0.06		23.2	1.96	0.15
	12.3	0.79	0.03		25.1	1.61	0.11
	15.2	0.69	0.04		27.5	1.34	0.08
	21.6	0.62	0.04		30.1	1.16	0.06
	25.5	0.57	0.03				

In fig. 2 for helium, the dotted curves refer to the theoretical values for scattering, assuming a Coulomb interaction between the nuclei and making allowance for the identity of the scattered and projected particles (Mott 1930). The experimental quantities seem to approach these values as the

TABLE II—THE RATIO OF OBSERVED TO CLASSICAL SCATTERING IN HYDROGEN

χ	Range mm.	" R "	P.E.
15°	10.6	2.33	0.25
	13.4	3.36	0.31
	18.8	7.0	0.6
	19.3	6.6	0.6
	23.8	11.4	1.1
21°	12.1	1.69	0.09
	14.9	2.45	0.14
	21.6	9.5	0.5
	25.7	13.0	0.7
33°	13.5	1.30	0.05
	17.5	1.58	0.09
	23.3	3.57	0.14
	27.4	4.47	0.16
45°	21.7	0.88	0.08
	26.7	0.58	0.06

FIG. 2—Anomalous scattering of α -particles in helium.
--- values calculated from Mott's formula.

velocity of the incident particles decreases, but perhaps not actually to reach them even at the least velocities available. The full curves denote theoretical values which are obtained by the use of appropriate adjustable constants which appear in relation (1) of § 4.

TABLE III—THE RATIO OF OBSERVED TO CLASSICAL
SCATTERING IN DEUTERIUM

χ	Range mm.	" R "	P.E.
15°	8.8	4.5	0.8
	11.5	24.9	2.6
	16.6	26.8	3.3
	20.3	35.5	3.2
21°	10.7	5.9	0.5
	13.1	14.8	1.2
	18.6	17.2	1.6
	23.6	20.0	1.7
33°	12.0	4.7	0.3
	14.9	6.1	0.5
	21.4	7.6	0.5
	25.5	16.8	0.9
45°	21.7	3.7	0.3
	26.7	4.9	0.4

The experimental values for helium at the higher ranges near 3 cm. agree fairly well with the earlier results, but at lower ranges there is a noticeable difference, and the values of R are consistently smaller. At 45° scattering we found a maximum at 2 cm. range, although less pronounced than that given by Chadwick. This was the most difficult of the observations because there were few scattered particles.

For the other two gases, the "Coulomb" scattering is simpler and corresponds to the value $R = 1$. As the range is reduced, measurement becomes less precise, so that it is not easy to find where the scattering ceases to be anomalous. We may conclude, however, that in all three cases it is slightly anomalous at 1 cm. range.

The variation of R with range and with angle for hydrogen (fig. 3) is similar in form to that found by Chadwick and Bieler (1921), but our absolute values are generally greater. The explanation may be that Chadwick and Bieler calculated the magnitude of the "classical scattering" from the dimensions of the apparatus: this procedure is likely to lead to inaccuracy.

The curves for deuterium (fig. 4) are different in form from those for hydrogen. The actual number of recoil nuclei collected in the counter for a certain region of velocities and angles is of the same order for the two gases

when these are used in the scattering chamber at the same pressure. This is in general agreement with the results of Rutherford and Kempton for the recoil nuclei projected in the forward direction. Allowance has then to be made for the difference in reduced mass of the α -particle and struck nucleus in the two cases, so that the values of R are different. A significantly greater resemblance is seen if these values are plotted for deuterium on a velocity scale extended in the ratio of the reduced masses, so that points with the same de Broglie wave-length are readily compared.

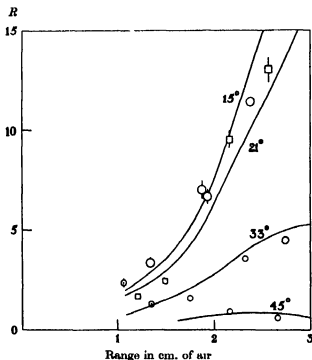


FIG. 3—Anomalous scattering of α -particles in hydrogen. □ 21°, ○ other angles.

Pollard and Margenau (1935) give results for protons and deuterons projected in the forward direction. They had no satisfactory method of obtaining the magnitude of the "classical scattering", and deduced it by assuming the yield to be classical at the lower ranges (1.4–1.7 cm. for hydrogen and 1.0–1.2 cm. for deuterium) and extrapolating to give the classical yield at higher energies. We find, however, that the scattering diverges considerably from the classical values at those ranges.

The yield of recoil deuterons obtained by Pollard and Margenau, when plotted against the range of the incident α -particles, has at about 1.4 cm. range an irregularity which becomes less obvious if R is plotted instead of

yield. The curve then becomes similar in form to those of fig. 4. Nevertheless, if this anomaly is real it seems to occur at a different range of the impinging α -particle as the angle of scattering is varied. This might preclude its explanation as a resonance effect.* The angular distribution appears to require rather the use of higher phase constants than the rapid variation with range of only one of them.

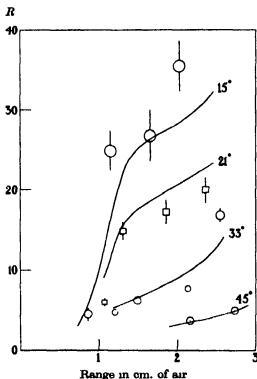


FIG. 4—Anomalous scattering of α -particles in deuterium.

The curves in figs. 3 and 4 denote theoretical values for hydrogen and deuterium, obtained in the same way as for helium.

4—DISCUSSION

If R denotes the ratio of the number of scattered particles to the number expected on classical theory, θ the angle of scattering with reference to axes

* Pollard and Margenau give 2.6 mV as the energy of the incident α -particles at which "resonance" occurs. When account is taken of the recoil of the deuteron, only one-third of this energy, 0.9 mV, is effective in the collision. The "breadth" of the resonance level would perhaps be the least attractive feature of the supposition.

moving with the centre of mass of the two particles, then (Mott and Massey 1933, p. 275)

$$R = |1 + A(\theta)|^2,$$

where

$$A(\theta) = iak \sin^2 \frac{1}{2}\theta \exp(i\alpha \log \sin^2 \frac{1}{2}\theta) \left\{ (e^{2iK_1} - 1) + \frac{3(1+i\alpha)}{(1-i\alpha)} (e^{2iK_1} - 1) \cos \theta \right. \\ \left. + 5 \frac{(1+i\alpha)(2+i\alpha)}{(1-i\alpha)(2-i\alpha)} (e^{2iK_1} - 1) P_2(\cos \theta) + \dots \right\}; \quad (1)$$

$$k = 2\pi M^* v / h; \quad ak = \frac{h\nu}{4\pi Ze^2}; \quad \alpha = 1/ak.$$

v is the velocity of the incident α -particles and M^* the reduced mass of the colliding particles. The quantities K are certain functions of v and the scattering potential which are discussed below.

For helium, if ϕ is the angle of scattering observed in the experiments,

$$\phi = \frac{1}{2}\theta, \quad (2)$$

and we have to modify the formulae to allow for the similarity of the scattered and struck nuclei. The formula for R will then not contain terms involving the odd parameters K_1, K_3 , etc. For hydrogen and deuterium, if χ is the angle of projection of the recoil nucleus,

$$\chi = \frac{1}{2}(\pi - \theta). \quad (3)$$

The quantities K which occur in formula (1) are defined as follows. Let $V(r)$ be the interaction energy between the colliding nuclei when at distance r apart so that $V(r)$ tends asymptotically to the form $2Ze^2/r$, where Z is the atomic number of the struck nucleus. Then the wave equation for the relative motion of the nuclei when the system possesses n quanta of angular momenta about the mass centre is

$$\frac{d^2 y_n}{dr^2} + \left[k^2 - \frac{8\pi^2 M^*}{h^2} V(r) - \frac{n(n+1)}{r^2} \right] y_n = 0, \quad (4)$$

where y_n is r times the wave function. The proper solution of this equation (i.e. the one which vanishes at the origin) will have the asymptotic form

$$\sin(kr - \alpha \log 2kr - \frac{1}{2}n\pi + \delta_n).$$

δ_n is thus the phase shift produced by the field $V(r)$. We now write δ_n in the form

$$\delta_n = \eta_n + K_n,$$

where $\eta_n = \arg \Gamma(n+1+i\alpha)$, is the phase shift which would be produced by the Coulomb potential $2Ze^2/r$ alone. K_n is then the phase shift due to the "anomalous" field acting at close distances.

The convergence of the series in (1) arises from the increase in the classical closest distance of approach between the nuclei as the angular momentum of the relative motion increases.

In analysing a given set of experimental results the first procedure is to determine at each velocity the phase parameters K_0, K_1, \dots , which, when substituted in (1), reproduce the observed angular distributions. In doing this, of course, as few terms of the series as possible are used. This procedure leads to the phase parameters given in fig. 5 as functions of the velocity. It will be seen that phases up to K_4 are important in helium, K_1 in hydrogen and K_2 in deuterium. These results are quite definite for, if an attempt is made to get the experimental curves with even one fewer term, angular distributions are obtained which, at one angle or another, deviate from the

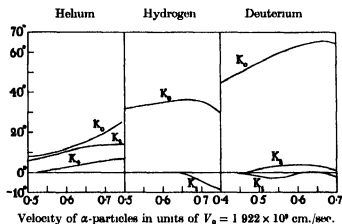


FIG. 5—Phase constants K_n .

experimental curves by amounts several times the probable error of the experimental readings. Moreover, the determination is unique (apart from the possible addition of an integral multiple of π) except for deuterium. For this case the number of observations for lower range α -particles is insufficient to distinguish between two possible sets of phase curves.

Having derived the phase parameters from the experimental results, the next procedure is to consider what information may be derived from the phases concerning the "anomalous" nuclear field. It is doubtful to what extent the nuclear interaction may be regarded as a static one. The probable existence of exchange types of force besides the likelihood of particle exchange between the colliding nuclei makes it unlikely that the interaction can be represented by a potential function depending only on the distances between the centres of mass of the two nuclei. Nevertheless, one very

general result follows quite definitely from the values of the phases derived in the manner described above. It has been pointed out already that a phase parameter K_n is only appreciable when the closest distance of approach of two nuclei possessing n quanta of angular momenta about their centre of mass falls inside the region where the Coulomb law of force breaks down. In fact, if m is the largest value of n for which K_n is important, the mutual separation of the nuclei at which deviations from the inverse square law first occur is given approximately by the root of the equation

$$k^2 - \frac{8\pi^2 M^* 2Ze^2}{h^2} \frac{m(m+1)}{r_0^2} = 0. \quad (5)$$

From this equation and the "observed" phases we find that the Coulomb law must break down at quite large distances, as much as 10×10^{-12} cm. or even greater. This is much larger than previously assumed, and invalidates the theoretical conclusions of Taylor for helium and hydrogen and of Pollard and Margenau for deuterium, as they considered only K_0 to be important.

The smaller order phases are largely determined by the interaction at close distances and we can examine the extent to which this interaction may be treated statically by representing the nuclear potential schematically in the form

$$\begin{aligned} V(r) &= 2Ze^2/r, & r > r_0, \\ &= -d, & r < r_0. \end{aligned} \quad (6)$$

This model may even be improved by breaking up the range of r for $r < r_0$ into two or more parts in each of which V has a different constant value. The parameters K for such a potential function may be calculated by methods described in the appendix. The determination of the best set of quantities r_0, d is by no means unique. Thus the potential at close distances need only be such as to give the correct value of K_0 , for the higher order phases are hardly affected by changes in V for small r . A narrow deep potential hole or a wider but shallower one near the origin can both be adjusted to give the observed values of K_0 . This flexibility may be important in relating static fields derived in this way from scattering with those derived from a study of nuclear binding energies.† The interaction potentials which lead to phase parameters agreeing most satisfactorily with those deduced from the

† Thus to the extent to which α -particles may be treated as single entities in nuclei, Massey and Mohr (1935) showed that a deep but short-range attraction must exist between two α -particles to explain the difference in binding energy between Be^8 and C^{12} .

experiments are illustrated in fig. 6. In deriving the depth of the deep attractive potential hole at the closest distances, one must pay attention to the ambiguity present in the phases due to the integral multiple of π which is indeterminate from the experimental results. This indeterminacy is resolved if we know the energy levels of the compound nucleus formed by

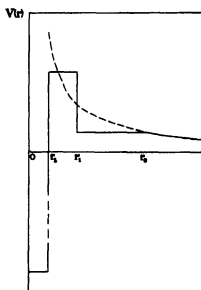


FIG. 6—Form assumed for the potential energy.

TABLE OF VALUES

Potential energies in mV
Radii in nuclear units of 10^{-12} cm

H_0	${}^1\text{H}$	${}^2\text{H}$
-39 $r < 1$	-130 $r < 1$	-41 $r < 1$
+3 $1 < r < 3$	-25 $1 < r < 3$	+1.4 $1 < r < 7$
+0.4 $3 < r < 14$	+0.3 $3 < r < 10$	+0.3 $7 < r < 10$
1s level -1.7	-35 (forbidden)	-0.1
1p level	-1	

the two colliding nuclei. If there is no energy level (even ignoring the Pauli Principle), then the multiple of π is zero; if a 1s level exists but no other, then $K_0 \rightarrow \pi$ as the energy of relative motion tends to zero; if 1s and 1p alone exist, both K_0 and K_1 tend to π and so on. In all cases it was assumed that a 1s level exists but that a 1p level occurs only for the α -p interaction. The very deep 1s level (-35 mV) which then appears in this case is excluded by the Pauli Principle as the 3 protons of Li^5 could not all occupy a 1s level,

and is not in contradiction to the low (even if positive) binding energy which this nucleus must possess.

The potential functions represented in fig. 6 do yield phases K which are in fair agreement with the "observed" phases, but there is no doubt that the observed variation of the phases with velocity is not accurately reproduced. This may be due to lack of precision in the form taken for $V(r)$ but is much more likely to arise from exchange forces which are known to have the effect of introducing, for the scattering of neutrons by heavy nuclei, an effective potential depending on the velocity (van Vleck 1935). We would like to make it clear, however, that, though it is doubtful what significance can be attached to the exact form of the internal potential function derived as above, the experimental results do lead to the determination of sets of phase parameters which do not depend on any theory of nuclear constitution and can be used to test any such theory which may be advanced.* The rather unexpected result that the range of the internal nuclear force is as great as 10^{-12} cm † is also free from any special hypothesis of nuclear interaction.

We wish to acknowledge gratefully our indebtedness to Lord Rutherford, Dr Chadwick and Dr Ellis for their encouragement and interest, and to Dr Massey for helpful discussions. One of us (G. E. P.) wishes also to acknowledge generous financial assistance from Emmanuel College and from the Department of Scientific and Industrial Research.

APPENDIX

1—CALCULATION OF WAVE FUNCTIONS FOR MOTION IN A COULOMB FIELD

Since it is frequently necessary to calculate Coulomb wave functions in a number of wave mechanical problems, we shall describe simple methods of obtaining them. We require the two independent solutions y_1, y_2 of the equation

$$\frac{d^2y}{dx^2} + \left[1 - \frac{2\alpha}{x} - \frac{n(n+1)}{x^2} \right] y = 0, \quad (7)$$

* Dr J. A. Wheeler has been using the available experimental material on the scattering of fast α -particles in helium as well as that for slow α -particles, in a thorough investigation of the interaction. We are indebted to him for much correspondence and for allowing us to make use of a table of his phase constants. Dr Wheeler's paper is to appear in the *Physical Review*.

† This result is not unexpected for deuterium, for the low binding energy of the H^2 nucleus means that it is much larger than more stable nuclei.

which are respectively zero and infinite at the origin, and which have at infinity the asymptotic form

$$\left. \begin{aligned} y_1 &\sim \sin (x - \alpha \log 2x - \frac{1}{2}n\pi + \eta_n), \\ y_2 &\sim \cos \end{aligned} \right\} \quad (8)$$

in which x is written for kr .

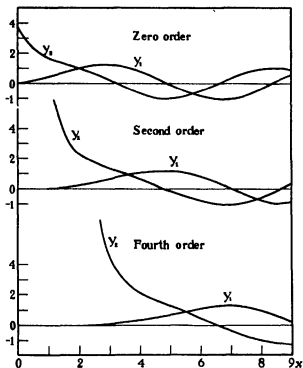


FIG 7—Wave functions of order 0, 2, 4, for relative motion of two α -particles, regarded as point masses with a Coulomb repulsion.

The usual method (Gordon 1928) of obtaining the solutions requires the calculation of a series which involves complex terms. It is easier, however, to use the expansions in real series to a distance at which the convergence becomes poor, and then to continue the solutions by means of several terms of the asymptotic expansion, for as soon as the convergence of the series expansion becomes poor, the asymptotic expansion is sufficiently good an approximation. For purposes of illustration, fig. 7 shows the form of the zero, second and fourth order wave functions for the motion

of α -particles of velocity $0.7V_0$ in helium (V_0 denotes the velocity 1.922×10^9 cm./sec.). It may be shown that

$$y_1 = \frac{1}{2} C (2x)^{n+1} \left(1 + \frac{\alpha x}{n+1} + \dots \right), \quad (9)$$

$$y_2 = \frac{C^{-1}}{2n+1} (2x)^{-n} (1 + \dots) + y_1 (A \log x + B), \quad (10)$$

where
$$C = \frac{(e^{2\pi\alpha} - 1)^{-\frac{1}{2}}}{(2n+1)!} \{2\pi\alpha(1+\alpha^2) \dots (n^2+\alpha^2)\}^{\frac{1}{2}}.$$

The series in brackets in (9) and (10) are the usual real series obtained by the method of Frobenius for the case in which the roots of the indicial equation differ by an integer. To find B we may evaluate the residues of the Mellin-Barnes integral for $W_{\alpha, n+\frac{1}{2}}(2ix)$ (Whittaker and Watson, 1927, p. 343 *et seq.*). The series so obtained, multiplied by $\Gamma(n+1-\alpha)$, has the desired asymptotic form, and B is obtained from the coefficient of x^{n+1} .

The same results are obtained by a substantially similar method by Yost, Breit and Wheeler (1936).*

The asymptotic expansions can be shown to take the form

$$\begin{cases} y_1 \sim U \sin \theta + V \cos \theta, \\ y_2 \sim U \cos \theta - V \sin \theta, \end{cases} \quad (11)$$

where $\theta = x - \alpha \log 2x - \frac{1}{2}n\pi + \eta_n, \quad (12)$

and U, V are the real and imaginary parts of the function

$$G(-n+i\alpha, n+1+i\alpha; 2ix),$$

where
$$G(\beta, \gamma, z) = 1 + \frac{\beta \cdot \gamma}{z \cdot 1} + \frac{\beta(\beta+1) \gamma(\gamma+1)}{z^2 \cdot 2!} + \dots \quad (13)$$

Very few terms of the series are required even for values of r near the first maximum of y . There is little difficulty in calculating the series (9) for y_1 so far as its first maximum, so that y_1 can be calculated readily for all required values of r .

In cases where it is laborious to calculate y_2 as far as its first zero, one may calculate it for intermediate values quickly by using the equation

$$\frac{y_2(r_2)}{y_1(r_2)} - \frac{y_2(r_1)}{y_1(r_1)} = \int_{r_1}^{r_2} \frac{1}{y_1^2(r)} dr. \quad (14)$$

If we know y_1 for all values of r and y_2 at any particular value—from the

* Further details are also given there. For useful tables compiled by these authors see 1935.

asymptotic expansion where this is valid— y_2 may be calculated approximately at any other value by simple numerical integration.

In concluding this section, it may be remarked that the quantity η_n appearing in the formula (12) may be most conveniently calculated from the equation

$$\left. \begin{aligned} \eta_n &= \eta_0 + \sum_1^n \arctan \frac{\alpha}{n}, \\ \eta_0 &= -0.5772\alpha + 0.4007\alpha^2 - 0.2074\alpha^3 + 0.1440\alpha^4 \\ &\quad - 0.1113\alpha^5 + 0.0910\alpha^6 + \dots \end{aligned} \right\} \quad (15)$$

2—CALCULATION OF THE NUCLEAR INTERACTION FROM THE PHASE CONSTANTS

Consider the solution of the wave equation (4) for the potential (6) of Gamow type. For the n th order wave, if $r > r_0$

$$y_n = y_1 \cos K_n + y_2 \sin K_n, \quad (16)$$

where K_n is chosen to give the asymptotic form

$$y_n \sim \sin(kr - \alpha \log 2kr - \frac{1}{2}n\pi + \eta_n + K_n). \quad (17)$$

Also if we put

$$\mu^2 = k^2 + \frac{8\pi^2 M^*}{h^2} d,$$

we have, for $r < r_0$

$$y_n = (\frac{1}{2}\pi\mu r)^{\frac{1}{2}} J_{n+\frac{1}{2}}(\mu r), \quad (18)$$

for this is the solution which vanishes at the origin.

The condition must now be satisfied that y_n and $\frac{dy_n}{dr}$ are continuous across the boundary at $r = r_0$. Elimination of the arbitrary constant relating (16) and (18) gives a relation between K_n and μ , and hence between K_n and d , the "depth".

In order to obtain a potential which gives the known values for the phases K_n at different velocities, one may readily generalize the method to a potential of the form:

$$V = -d_1, \quad r < r_1,$$

$$V = -d_2, \quad r_1 < r < r_2,$$

and so on, where r_1, r_2, \dots are less than r_0 .

In using this method one makes use first of the velocity-variation of the highest important phase to fix r_0 . The lower phases determine progressively those parts of the field nearer the origin. For higher orders, and small values of r, y_1 and its derivative are small, y_2 and its derivative are large. This has

the effect of making the corresponding phases practically independent of the field at those values of r . y_1 and y_2 become comparable at the point of inflexion given by the equation (5).

SUMMARY

The scattering of α -particles of polonium has been observed in helium, hydrogen, and deuterium over a region of different velocities and scattering angles.

The scattering associated with Coulomb fields takes place only at ranges much lower than was previously thought. This indicates departure from the Coulomb law of force at distances greater than 10^{-12} cm.

The ratio of observed to "classical" scattering was found to be large for hydrogen and deuterium, but the results for helium were found generally to be lower than previous estimates. The results are interpreted in terms of a nuclear potential of the Gamow type.

The resulting nuclear potentials are discussed. They can be made to conform with data on the binding energies of nuclei.

REFERENCES

- Chadwick 1930 *Proc. Roy. Soc. A*, **128**, 114.
Chadwick and Bieler 1921 *Phil. Mag.* **42**, 923.
Gordon 1928 *Z. Phys.* **48**, 180.
Massey and Mohr 1935 *Nature, Lond.*, **136**, 141.
Mott 1930 *Proc. Roy. Soc. A*, **126**, 259.
Mott and Massey 1933 "The Theory of Atomic Collisions." Oxford.
Pollard and Margenau 1935 *Phys. Rev.* **47**, 833.
Rutherford and Chadwick 1927 *Phil. Mag.* **4**, 605.
Rutherford and Kempton 1934 *Proc. Roy. Soc. A*, **143**, 724.
Taylor 1931 *Proc. Roy. Soc. A*, **134**, 103.
— 1932 *Proc. Roy. Soc. A*, **136**, 605.
van Vleck 1935 *Phys. Rev.* **48**, 367.
Whittaker and Watson 1927 "Modern Analysis."
Wright 1932 *Proc. Roy. Soc. A*, **137**, 877.
Wynn-Williams and Ward 1931 *Proc. Roy. Soc. A*, **131**, 391.
Yost, Breit and Wheeler 1935 *Terr. Magn. Atmos. Elect.* **40**, 442.
— — — 1936 *Phys. Rev.* **49**, 2, 174.
-

The Magneto-Resistance effect in Cadmium at Low Temperatures

By C. J. MILNER, B.A.

Hutchinson Research Student, St John's College, Cambridge

(Communicated by Lord Rutherford, O.M., F.R.S.—

Received 29 December 1936)

INTRODUCTION

The experiments of Kapitza (1929) showed that the increase of electrical resistance produced in a metal by a magnetic field H is not proportional to H^2 , as was previously supposed. In the new experimental range made available by his method (Kapitza 1927) of producing very strong fields up to 300 kilogauss, Kapitza found that the increase of resistance tended towards a linear variation with the field strength. The result may be expressed in the formula

$$\Delta R/R_0 = b(H - H_k), \text{ for } H \gg H_k,$$

where R_0 is the resistance at 0° C. This gives the asymptote to the experimental curve: but if experiments are made at field strengths up to a maximum H_m , and $H_m \gg H_k$, then over a large part of the experimental range the curve obtained is practically identical with the asymptote. If the linear part of the curve is then extrapolated back to meet the axis of H , its intercept on that axis gives the parameter H_k , and the slope of the line gives the parameter b . If, however, the maximum field used is only of the order of H_k , the linear variation is only reached outside the experimental range; and some formula must be employed, in effect, to extrapolate to the region where the linear law holds, before the position of the asymptote and the values of the parameters can be derived. It is obvious that the values so obtained will vary according to the particular formula adopted.

Kapitza found that the linear variation was a general phenomenon shown by all metals; and this result has been confirmed (in cases where a sufficiently strong field has been available) by the experiments of Meissner and Scheffers (1929, 1930), of de Haas and van Alphen (1933), and of de Haas and Blom (1934, 1935). Kapitza also found, in studying the parameters obtained for different specimens of a metal at different temperatures, that for the "true metals" the "critical field" H_k decreased with decreasing temperature

(for the semi-conductors H_k was independent of temperature). The value of H_k was greatly increased if the specimen was strained or only slightly impure. The coefficient b (which corresponds to Kapitza's β_k) was independent of the purity of the specimen, and varied with the temperature alone. In the majority of "true metals" the critical field was too high for separation of the parameters to be possible at temperatures higher than that of liquid nitrogen, which was the lowest used in his research.

The initial object of the present work was to examine in more detail this observed dependence of the parameters on the purity of the specimen. Cadmium was selected for these experiments, since Kapitza showed that it had a low critical field. Thus his specimen Cd_{11} gave a value for H_k of only 12 kilogauss at liquid nitrogen temperature: and this field strength could be considerably exceeded by using the maximum field (26 kilogauss) of the electromagnet which was available for this work.

EXPERIMENTAL METHODS

Throughout this research the transverse magneto-resistance effect has been studied, except in one experiment in which the parallel effect was examined: the magnetic field was perpendicular to the current within 2 or 3° . The specimens were about 2 cm. long, and the magnetic field was constant to 1 % over the specimen. The form of the specimens used may be seen in fig. 1, where a specimen is shown in place in the Dewar flask which contained the liquefied gas. A loop of cadmium wire, of 0.35 mm. diameter, formed the resistance and its current leads, two potential leads of the same wire were fused to this at A and B with a pinhole gas flame. The four leads from the resistance ran in grooves cut in a quartz plate which supported and separated them. The joints with copper or manganin leads were grouped close together at J , to reduce thermoelectric effects.

The Dewar vessel was mounted between the poles of the electromagnet on a parallel-rule device, so that it could be swung clear of the field, and then replaced, as desired. In this way the resistance in a field H and that in zero field could be rapidly compared. A search coil fixed to the outside of the Dewar flask measured the field strength as the flask was removed from or replaced in the field, by the deflexion of a fluxmeter to which it was connected.

The measurements of resistance were made with a potentiometer arrangement, slightly modified to allow a high sensitivity to be used for the measurement of the very small resistances of such specimens. These modifications have been described elsewhere (Kapitza and Milner 1937). A galvanometer

combination following the lines discussed by Hill (1934) was used: the primary instrument, a Broca astatic, had a sensitivity of about 40 mm./ μ V (for 1 m. scale distance), and the photoelectric magnification produced an overall sensitivity of about 250 mm./ μ V. This seemed to be the "useful limit" in these experiments; though more amplification could easily have been obtained, the unsteadiness began to be appreciable at this value. An advantage of the system was that the primary galvanometer (only *nominally* astatic) could be placed in another room where the stray field of the electromagnet had only a slight effect on its zero.

An important source of unsteadiness was eliminated by compensating the e.m.f.'s induced in the loop of the potential leads to the specimen by fluctuations in the magnetic field. A second coil wound on the outside of the Dewar flask was connected to the galvanometer circuit through a potential divider and reversing switch. The potential divider was adjusted so that the e.m.f.'s in the circuit induced in the loop of the leads to the specimen were compensated by equal and opposite e.m.f.'s from the coil.

When this was done, the remaining fluctuations were not affected by switching on the measuring current in the resistance. Consequently, the variations of resistance due to the varying field did not appear (as was suggested by de Haas and van Alphen 1933) to be an important source of fluctuations, even when, as in this work, the electromagnet was supplied from a dynamo, and not from batteries.

From time to time during each experiment a test was made to see that the ratio of the currents in the specimen and the potentiometer remained constant by balancing on the potentiometer the potential across a standard 10 Ω coil connected in series with the specimen. A normal potentiometer circuit was used for this (breaking the galvanometer circuit to test for balance). Fig. 2 shows the electrical circuit employed, and the switching by which the balancing was arranged. *G* is the primary galvanometer

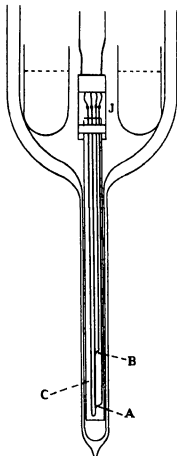


FIG. 1—Specimen in Dewar flask.

(Broca); S is the secondary instrument (an ordinary moving-coil); and C is the copper-copper oxide differential photocell.

A series of ten or twelve observations of the resistance were made at fields increasing each time by about 2 kilogauss up to 26 kilogauss. The resistance measurement was repeated for the maximum field with the field reversed, to test for the absence of a Hall potential.

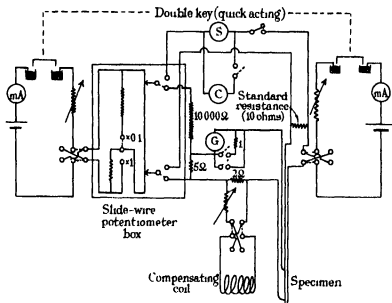


FIG. 2—Circuit used for magneto-resistance measurements.

EXPERIMENTS AT LIQUID HYDROGEN TEMPERATURE

It was decided to study the dependence of the magneto-resistance effect on the degree of purity in some detail at the temperature of liquid hydrogen, since this was easily available, and at this low temperature the critical fields were only of the order of 3 kilogauss. The latter fact allowed the linear variation of resistance to be fully established in the experimental range of field strengths.

A series of cadmium specimens was prepared of varying purity and extent of annealing. The sources and treatment of the specimens are given in Table I, together with the ratio R/R_0 of the resistance at 20.4°K. to that at the ice-point. The Hilger cadmium (specimens A, B, and C) was very pure, Messrs Hilger's spectroscopic analysis showing lead as the only impurity, the amount being about 1 part in 10^6 .

The Kahlbaum specimen D was found by spectroscopic comparison to be of approximately equal purity. The Merck cadmium contained about 0.01 % each of Cu, Ag, and Pb, again estimated spectroscopically. The alloys with 0.1 and 0.4 % Cu were prepared by melting in vacuo a weighed quantity of the Merck cadmium and dissolving in this a weighed length of copper wire. The mixture was heated to 400–450° C. and shaken for 10 min. before being allowed to cool and solidify.

TABLE I

Ident. letter	Source	Treatment	$10^4 \times R/R_0$
A	Hulger	15 hr. at 300° C.	207.0
B	Hulger	As extruded	209.2
C	Hulger	3 hr. at 220° C.	208.0
D	Kahlbaum	Extruded by Kapitza (1929)	206.0
F	Merck	6 hr. at 220° C.	224.5
G	Merck + 0.1 % Cu	3.5 weeks at room temperature	282.2
H	As G	+ 3.5 hr. at 220° C.	280.2
J	Merck + 0.4 % Cu	As extruded	347.4
K	As J	+ 16 days at room temperature	324.0
L	As D	+ 5 months at room temperature	203.6
M	As L	+ 1 month at room temperature	203.3

In each case the material was obtained in or machined into the form of rods 5 mm. in diameter. These were extruded into wires of 0.35 mm. diameter through a steel die, using a hydraulic press; a pressure of 5000 atmospheres at 150° C. was suitable. Spectroscopic comparison showed that the material was not contaminated during these processes.

It was found that the specimens became annealed at room temperature. this was judged from the fact that the resistance of a specimen continually fell for some time after preparation, even when no heat treatment had been applied. The time since preparation has therefore been included in the Table, as well as the extent of the high-temperature annealing which was given. That annealing at room temperature can be fairly complete may be judged from the behaviour of the specimen G. This was allowed to stand for 3½ weeks, during which the resistance ratio dropped from 308 to 282. Further annealing for as much as 3.5 hr. at 220° C. only caused a further drop to 280. Owing to this effect, and to the reverse action of recent straining by cooling, we considered each experiment as made on a fresh specimen.

In order to correct the measured resistances for the variation of the boiling point of hydrogen with barometric pressure, we first obtained the temperature of the experiment from tables of the vapour pressure. The curves published by de Haas, de Boer, and van den Berg (1935) were then

used to obtain the variation of the resistance ratio of cadmium with temperature near to the boiling-point of hydrogen, and the correction was added to or subtracted from the measured resistance ratio accordingly.

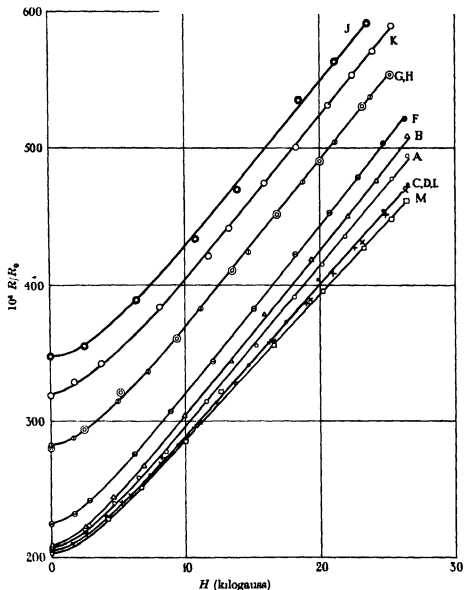


FIG. 3.—Magneto-resistance curves at liquid hydrogen temperature— R/R_0 against H .

We did not measure the ice-point resistance directly in each case. The resistance of one of the specimens, B, was determined at a series of temperatures between 0 and 30° C., and a correction curve was prepared from

these data, so that the ice-point resistance could be obtained from the resistance at the room temperature. This resistance was found to be proportional to the absolute temperature within $\frac{1}{4}\%$; and it was assumed that the same proportionality would hold for all the specimens. The correction was between 6 and 10 % in all cases.

The change of resistance was examined in the manner already described. In fig. 3 the values of the ratio R/R_0 obtained for the various specimens are plotted against the corresponding field strengths. The asymptote to each curve was independently estimated and the parameters deduced are plotted against the resistance ratio (for zero field), H_k in fig. 4 and b in fig. 5

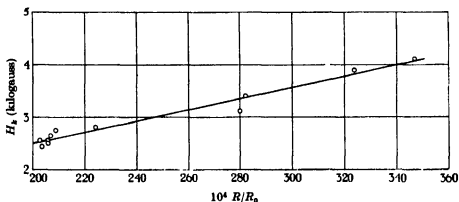


FIG. 4— H_k against R/R_0 , for the curves of fig. 3.

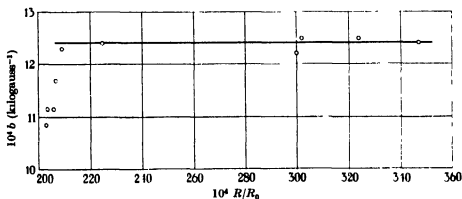


FIG. 5— b against R/R_0 , for the curves of fig. 3.

DISCUSSION OF RESULTS AT LIQUID HYDROGEN TEMPERATURE

Considering first the graph of fig. 4, it will be seen that the critical field H_k appears to depend only on the resistance ratio, within the accuracy of estimation. The variation appears to be linear over the range examined:

the best straight line passes very close to the origin ($H_k = 0.3$ kilogauss at $R/R_0 = 0$), so that the critical field is approximately proportional to the resistance ratio, i.e. to the specific resistance.

This result is of interest as confirming a theoretical suggestion of Peierls (1931). He supposes that the quadratic law of change of resistance, which must hold in sufficiently low fields, might well break down when the paths of the electrons in the metal are bent through a considerable angle by the magnetic field during the mean time between collisions of the electrons with the atoms of the lattice. Expressing the condition that the radius of the path is of the order of the mean free path, and substituting for the velocity of the electrons the value obtained from the Fermi statistics, and for the mean free path the value derived from the conductivity, Peierls obtains for the critical field

$$H_c = ecn\rho,$$

where n = number of electrons per c.c., ρ = specific resistance, and e and c are, as usual, the electronic charge and the velocity of light. Since $R/R_0 = \rho/\rho_0$, the number n may be calculated from the slope of the graph of fig. 4, and the value may be expressed as 0.021 electron per atom, using known values of density, atomic weight, and other constants. It should be observed that our H_k and Peierls's H_c are not defined in the same way, and that they might well be proportional without being equal: only an agreement in order of magnitude can be expected.

Kapitza (1929) found, however, from his experiments at the temperature of liquid nitrogen that H_k varied much more rapidly than proportionally with the ratio R/R_0 : in fact, that H_k was approximately proportional to the additional resistance defined by the Matthiessen-Nernst rule, i.e. the residual resistance. This quantity may be obtained for our specimens by subtracting from the measured R/R_0 the value of the ideal resistance (which is by definition independent of the purity) at the temperature of the experiments 20.4° K. This has been measured by Meissner and Voigt (1930), who found $10^4 \times R_i/R_0 = 202.0$, and by de Haas, de Boer, and van den Berg (1935), who also proved in their experiments that the application of the Matthiessen-Nernst rule to cadmium was valid. Their value was $10^4 \times R_i/R_0 = 197.5$. We measured the residual resistances of two of our specimens, B and C, and obtained the value 199 for $10^4 \times R_i/R_0$. Subtracting this from the values of R/R_0 set out in Table I, we obtain values of the additional resistance ranging from 7×10^{-4} to 140×10^{-4} , i.e. varying by a factor of 20. The critical field, however, did not vary (see fig. 4) by so much as a factor of 2. We see therefore that the rule enunciated by Kapitza cannot hold for our case of cadmium at the temperature of liquid hydrogen.

Turning now to fig. 5, it will be seen that over almost all the range of R/R_0 the value of b is constant at $10^4 \times b = 12.5$ kilogauss $^{-1}$. The most perfect specimens, however, show values of b which are up to 10 % lower than this. The cause of these lowered values is suggested by the observation that considerable decreases, more than the experimental error, only appear for specimens which are both very pure and very well annealed. In these cases extensive recrystallization had probably taken place, with a consequent growth of "preferred" orientations among the crystallites at the expense of less preferred ones. The resulting wire would not therefore be a true polycrystal, in the sense that the condition of random orientation would not obtain. Some experiments with cadmium single crystals (Milner 1937) showed that the parameters of the magneto-resistance effect are both affected by crystal orientation, but especially the value of b is so dependent: we suggest that the change in orientation of the crystallites is the probable explanation of the altered values of b for the annealed wires. The occurrence of considerable recrystallization was demonstrated by etching the Kahlbaum specimen, which showed the lowest value of b . The average grain size was about 0.1 mm.: a freshly extruded wire showed grains of only about 0.01 mm. in size. We suggest therefore that the value of b is probably constant for all genuinely polycrystalline wires. This result is in agreement with Kapitza's observations.

EXPERIMENTS AT LIQUID NITROGEN TEMPERATURE

The result obtained in the experiments at liquid hydrogen temperature; that the critical field varied proportionally with the total resistance, differed from Kapitza's observation that H_k was proportional to the residual resistance. We therefore decided to examine some specimens of widely differing additional resistance at the temperature of liquid nitrogen,* at which Kapitza's experiments were made.

We examined specimens made from (i) the very pure Hilger cadmium, (ii) the Merck cadmium, (iii) the "0.4 % alloy" (see Table I). The points obtained are plotted in fig. 6, which shows $\Delta R/R_0$ with H . They lie too close for separate curves to be drawn through them. We also tested the specimen of Kapitza's cadmium from Kahlbaum, and the points and curve obtained for this are also given in the figure. It will be seen that the change of re-

* Chemically purified nitrogen was used in these experiments. The importance of its use in magnetic experiments when a constant temperature is required has been emphasized elsewhere (Kapitza and Milner 1937). The liquid obtained from our source contained only approx. 0.3 % of oxygen.

sistance is the same within experimental error for all the specimens except the Kahlbaum: for this it is about 10 % less than for the others. This would correspond with the smaller value of b found for this specimen at hydrogen temperatures (see fig. 5).

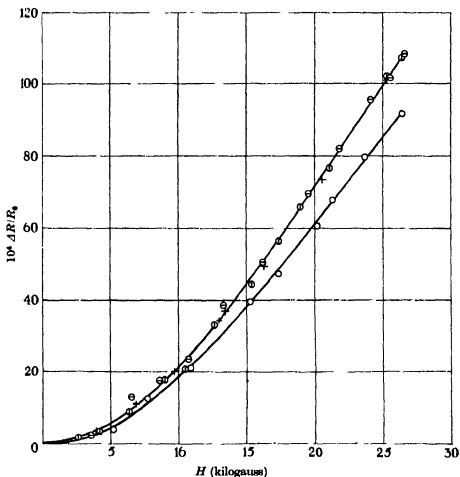


FIG. 6.—Magneto-resistance curves at the temperature of liquid nitrogen— $\Delta R/R_0$ against H .

The critical field is too high to admit of an accurate determination of the parameters H_k and b . It may, however, be roughly estimated that $H_k = 10$ kilogauss for all the specimens, and that $b = 6.5 \times 10^{-4}$ kilogauss $^{-1}$ for all except the Kahlbaum specimen. For this $b = 5.5 \times 10^{-4}$ kilogauss $^{-1}$ approximately.

The coincidence of the curves is again in agreement with the formula of Peierls: in these experiments the total resistance was the same for all the

specimens to about 1 %. The additional resistances varied again by a factor of about 20, on the other hand; and in Kapitza's view the change of resistance would have been altered proportionately.

EXPERIMENTS AT LIQUID HELIUM TEMPERATURES

With the development of the new helium liquefier constructed by Kapitza (1934) into a reliable and very efficient machine, the use of liquid helium has become a simple operation. It was therefore natural to examine the magneto-resistance effect in the new temperature range thus opened up. With the exception of the recent papers of Schubnikow, de Haas, and Blom (1935) on bismuth, and of de Haas and Blom (1935) on gallium, no experiments have so far been published in which the magneto-resistance effect has been examined at helium temperatures in fields greater than those obtainable with a solenoid, say 1 kilogauss. Attention has been mainly directed to experiments on supra-conductors, where interest lies especially in the part of the curve where the resistance is reappearing during the transition from the supra-conducting to the normal state.

It was realized that, contrary to normal practice, it was possible to use liquid helium in a single Dewar vessel, and that no jacket of liquid hydrogen or liquid nitrogen was essential. This made it unnecessary to alter the magnet poles and other experimental arrangements for these measurements, except that a new Dewar flask was made having a larger reservoir for the liquid (550 c.c. instead of 150 c.c.). This was necessary on account of the very low latent heat of liquid helium, which indeed is responsible for most of the complications of the technique.

Otherwise, all usual precautions were taken to reduce the heat entry into the flask. The rate of boiling with the flask "full", with the liquid helium level about 10 cm. below the cap, was about 250 l. (of helium at N.T.P.) per hour; and when the flask was nearly empty about 100 l. per hour. The 400 c.c. of liquid lasted about 1 hr. This rate of boiling was the maximum which could be allowed with any convenience.

The specimen of Kahlbaum cadmium "D" was first examined. This was very pure, having a residual resistance of only 5.5×10^{-4} . The first experiment gave a change of resistance much larger than was expected, and it appeared to be approximately proportional to H^2 instead of to H . When the experiment was repeated, the residual resistance was found to be slightly larger, 6.0×10^{-4} , and the curve found was somewhat lower. This is plotted in fig. 7 (curve 1). The pressure over the liquid helium was reduced, making the temperature 2.35° K., and the curve 2, fig. 7, was obtained, the increase

of resistance being some 10 % higher. Curves 3 and 4 show by comparison the magnitude of the effect at 20.4 and 14.2° K. respectively, for the same specimen.

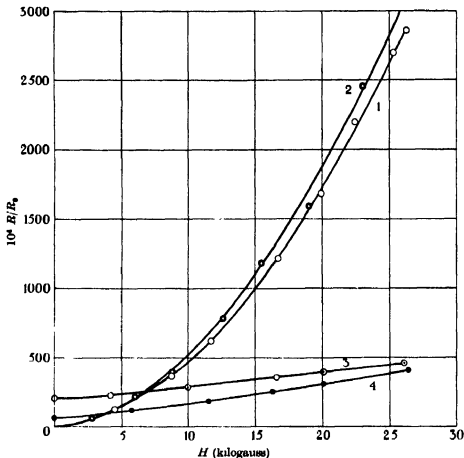


FIG. 7.—Magneto-resistance curves at liquid helium temperatures.

Other curves were obtained with (i) the same specimen further strained, residual resistance 6.5×10^{-4} , (ii) the Merck cadmium specimen "F", residual resistance 20×10^{-4} , (iii) the "0.4 % alloy", residual resistance 92.5×10^{-4} . In every case the results were consistent with the formula

$$\Delta R/R_0 = a.H^2 + b.H$$

and not with a simple square law. This is shown in fig. 8, where the results of the various runs are plotted as $(\Delta R/R_0)/H$ against H : so that the slope of the straight line is a , and the intercept on the vertical axis is b , in the above formula.

Within the experimental error, it may be seen from fig. 8 that b is constant for all the curves taken at 4.2° K. , the value being close to $b = 7 \times 10^{-4}$ kilogauss $^{-1}$. For the lower temperature, 2.35° K. , b is higher ($b = 10 \times 10^{-4}$ approx.). The values of a and b deduced for the various curves are set out in Table II.

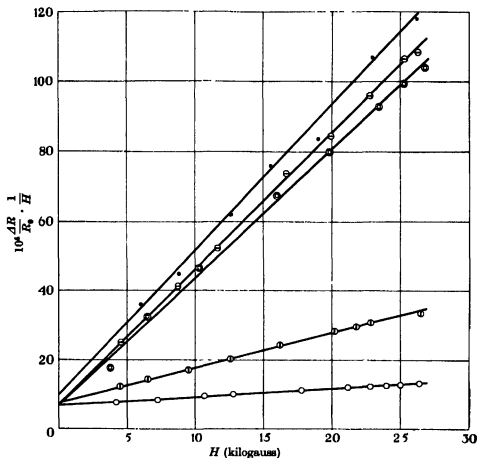


FIG. 8—Linear relation between $\left(\frac{\Delta R}{R_0} \cdot \frac{1}{H}\right)$ and H at the temperature of liquid helium.

TABLE II

T ($^\circ \text{ K.}$)	$10^4 \times R/R_0$	$10^4 \times a$	$10^4 \times b$	$10^4 \times a \times R/R_0$
4.2	6.0	3.88	7.5 (± 1)	23.5
4.2	6.5	3.65	7.0 (± 0.6)	22.1
4.2	20.0	1.01	7.7 (± 0.2)	20.2
4.2	92.5	0.24	7.05	22.2
2.35	6.0	4.2	10 (± 2)	25.2

DISCUSSION OF RESULTS AT LIQUID HELIUM TEMPERATURES

The fact that the coefficient b of the linear term is found to be constant at constant temperature supports the identification of this term with the linear or "Kapitza" effect found at higher temperatures, if we assume that the critical field H_k has decreased to negligible values at helium temperatures.

We thus have the following values of b for cadmium at various temperatures: at 2.35°K. , $10^4 \times b = 10$; at 4.2°K. , 7; at 20.4°K. , 12.5; and at 77°K. , 6.5 (kilogauss $^{-1}$). It would seem that the value of b must vary in some complicated manner, possibly periodically, with the temperature; but not very widely, probably not by much more than a factor of 2.

In contrast to the linear or " b effect", it would seem that the square law or " a effect" is strongly dependent on the purity of the specimen: the constancy of $a \times R/R_0$ in column 5 of Table II shows that the value of a is inversely proportional to the residual resistance.

DEPENDENCE OF THE a EFFECT ON TEMPERATURE

It is evident from the observations made at 2.35°K. that the a effect does not vary much with temperature below the boiling point of helium. At 20.4°K. , however, the graphs of ΔR with H tend to perfect linear asymptotes, and it would seem that at this and higher temperatures the a effect is not appreciable. It therefore becomes of interest to determine the manner of the variation of a with the temperature in the range between 4 and 20°K. Somewhat complicated apparatus is required if a constant temperature in this range is to be maintained by the use of a "helium vapour cryostat": it was decided to try to observe the variation of resistance with temperature in a constant magnetic field, while the specimen warmed up after the liquid helium had all evaporated.

A copper-constantan thermocouple was soldered to the point of the current lead adjacent to the centre of the specimen (at C in fig. 1) and its e.m.f. (against a second junction in ice-water) was used to determine the temperature of the specimen. The thermocouple was calibrated in a separate experiment, in which it was immersed in liquid hydrogen and in solid hydrogen. The deflexions of the galvanometer to which it was connected were obtained for these temperatures (20.4 and 14.2°K.), and a parabolic relation was assumed ($E = A + BT + CT^2$) between the deflexion and the absolute temperature. The thermoelectric power was of the order of $2 \mu\text{V}/^\circ\text{K.}$

over this range. It was verified that the e.m.f. did not depend on the magnetic field (we had reason to suspect that, with cadmium, the thermoelectric power did depend on the magnetic field at the junction).

The experiment was made with the Kahlbaum specimen "D", in which the α effect was most pronounced. It was found that the rate of warming-up was too rapid for more than one or two measurements to be made between 4 and 20° K. The rate of temperature rise was reduced in later experiments by means of the following device. An annular glass cup was placed in the reservoir of the Dewar flask (see fig. 1), and the flask was initially filled with liquid helium to cover the top of this cup. The liquid in the cup was effectively thermally insulated, except from heat radiated downwards from the cap of the vessel, by a "guard ring" of helium vapour, as long as any liquid remained in the appendix of the flask. As a result, there was still liquid helium remaining in the cup when all the liquid in the main part of the flask had boiled away: this liquid acted as a "source of cold" within the flask which reduced the rate of access of heat to the specimen, and thus the rate of temperature rise. The time taken for the specimen to reach 20° K. was raised in this way from about 1 min. to about 7 min.

The experiment was repeated using this arrangement, with field strengths 22.8, 10.5 kilogauss, and zero. The curves of R/R_0 with the absolute temperature T are given in fig. 9.

In these three curves we have theoretically sufficient data to evaluate α and β at any temperature, i.e. we have values of $\Delta R/R_0$ at two field strengths, but the accuracy of the experiments is not considered sufficient to justify this. However, to a first approximation we may treat the linear β effect as constant. On this assumption the difference in slope of the curves of fig. 9 represents the variation of the α effect with temperature.

It will be seen that above about 15° K. the differences in the ordinates of the curves are practically in the ratio of the field strengths (i.e. only the linear effect is present). The curves rise rapidly below this temperature, with a point of inflexion at about 6.5° K., below which the resistance appears to tend to a constant value at the lowest temperatures.

From these results it appears that the " α effect" is of a different origin from the " β effect", and that at these low temperatures a new mechanism of the increase of resistance comes into play. Such a new mechanism has been suggested by Peierls (1930), who showed that the quantization of the circular orbits of the electrons in a magnetic field may influence the resistance if μH (μ being Bohr's magneton) is comparable to kT . This condition is satisfied in our case, for at 10 kilogauss we have $\mu H \sim 10^{-16}$, while, at 4° K. $kT \sim 5.5 \times 10^{-16}$. This mechanism would be closely related to the

diamagnetism of conduction electrons, and one might expect it to be marked in a strongly diamagnetic substance like cadmium.

It would appear that the rapid decrease of the α effect with increasing temperature could be adequately accounted for by this mechanism: it would seem difficult, however, to understand the observed strong dependence of the effect on purity.

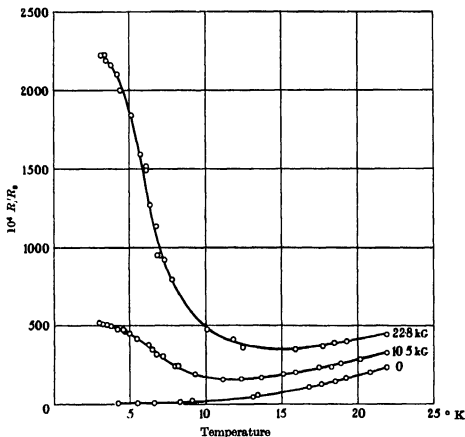


FIG. 9—Dependence of magneto-resistance on temperature— R/R_0 against T .

THE ABSENCE OF THE α EFFECT WHEN THE MAGNETIC FIELD IS PARALLEL TO THE CURRENT, AND THE SATURATION OF THE b EFFECT

It was established by Kapitza (1929) and also by the more detailed investigations of de Haas and van Alphen (1933) that the change of resistance with the magnetic field parallel to the current was of the same order as the perpendicular effect which we have discussed hitherto. For all the elements

examined, except gallium, but including cadmium, the "parallel" effect was between two-thirds and one-half of the perpendicular effect; and the curves for the variation of ΔR with H were similar for the two positions (i.e. the critical fields were nearly the same).

Since we considered that we had recognized a new effect (the a effect) appearing at helium temperatures in addition to the Kapitza or b effect, we decided to examine the parallel change of resistance at this temperature, to see if the ratio $\Delta R_{\parallel}/\Delta R_{\perp}$ remained unchanged. A specimen of the pure Hilger cadmium was prepared in the same way as previously, but only 6 mm. in length; and this was mounted on a strip of mica so that it was horizontal when placed in the appendix of the Dewar flask. It could then be turned about a vertical axis so as to move it from the parallel position to the perpendicular position. The specimen and its leads were fastened down to the mica so as to lie in a plane. This plane was then set parallel to the magnetic field by adjusting it so that the loop of the specimen and its potential leads included no lines of force. When this adjustment was correct, no "kick" of the galvanometer connected to it occurred when the field strength was altered.

The change of resistance in the parallel position was then determined at liquid helium temperature. It was found to be very much smaller than in the perpendicular position: for 25 kilogauss, $\Delta R/R_0$ was about 80×10^{-4} instead of about 2000×10^{-4} for the perpendicular position. The results are shown in fig. 10. It will be seen that the curve of $\Delta R/R_0$ rises linearly from the origin, and has a slight curvature downwards. The initial slope of the curve is approximately $10^4 \times b = 4$; i.e. about 0.6 of that found for the perpendicular effect at this temperature. De Haas and van Alphen found the ratio $\Delta R_{\parallel}/\Delta R_{\perp}$ for cadmium alloyed with 1 % of mercury in a field of 23.6 kilogauss to be 0.648 at (we infer) liquid hydrogen temperature. Kapitza found $\beta_{\parallel}/\beta_{\perp} = 0.83$, for pure cadmium at liquid nitrogen temperature. The agreement is as good as can be expected, considering the nature of the comparison, if we assume that the square-law a effect, which produces the large ΔR_{\perp} at helium temperatures, is not active in the parallel position. We are thus led to consider the a effect as probably quite different in origin from the b effect.

This experiment also strongly suggests that the increase of resistance due to the "Kapitza" or b effect reaches a saturation value for high fields and low temperatures, the saturation being masked by the a effect in the perpendicular case. We found that the method described above for setting the specimen to the parallel position was not very accurate. When the curve A of fig. 10 had been taken, a resetting of the specimen by the induction e.m.f.

method described was made; and the increased value of $\Delta R/R_0$, B in the figure, was obtained, with the maximum field applied. We then rotated the specimen to find the position at which the minimum value of ΔR actually occurred: this minimum value is also shown, C in the figure.

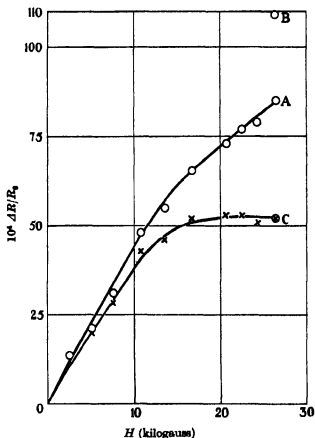


FIG. 10.—Parallel effect at the temperature of liquid helium— $\Delta R/R_0$ against H . Experimental values \circ ; derived values (see text) \times .

If we assume that the increases CA , CB are due to a small term $a' \times H^2$ resulting from the inaccuracy of the previous settings (i.e. that the current had a small component perpendicular to the magnetic field), and correct the other points accordingly by an amount proportional to H^2 , we obtain the revised points as shown, following the curve up to C . On this assumption the resistance is practically independent of the field beyond about 15 kilogauss.

Such behaviour of the Kapitza effect has not been noted previously,

though it has been predicted on theoretical grounds by Frank (1931), who obtained the equation

$$\Delta R/R_0 = BH^2/(1 + CH^2)$$

as representing the change of resistance. Assuming this formula, it is easy to show that the tendency to saturation should be evident at about $10 \times H_k$. Only very few experiments have been made in fields so much higher than H_k , and in which also the conditions have excluded the appearance of an a effect. (The co-existence of an effect of the Frank type and a square-law effect would obviously allow of the apparent linear region of the curve being greatly extended.) Since the same type of function has been found to represent both the parallel and the perpendicular effects at higher temperatures, we may suppose that the b term of the perpendicular effect at helium temperatures is also strictly to be represented by a saturating function. The difference between this and the linear term assumed in the equation

$$\Delta R/R_0 = a \times H^2 + b \times H$$

could not be seen, for example in fig. 9, owing to the superposition of the very large a effect in the perpendicular case.

COMPARISON WITH THE RESULTS OF OTHER EXPERIMENTERS

1—Kapitza (1929)

We confirm the existence of an initial square law and a subsequent approach to a linear law in the curve of ΔR with H : but we obtain some evidence of ultimate saturation. We confirm that the coefficient b of the linear law is independent of the perfection of the specimen. We find, however, that the parameter H_k is proportional to the total resistance, and not to the additional resistance, at hydrogen temperatures, and our measurements at nitrogen temperatures do not support the conclusions drawn by Kapitza from his results. If we examine, however, his fig. 12 (1929, p. 309), we see that the disagreement is not great, and that a good fit with his points could equally be made by assuming the same value of H_k for his two cadmium specimens, and different values of b , instead of vice versa. Thus b would be about 7% lower for his specimen Cd₁, which he remarks (1929, p. 312) to have been much softer than Cd₁₁. We have found the change of resistance, and b , to be about 10% lower for a very well annealed wire, probably owing to a preferred orientation among the crystallites.

His generalization, that H_k is proportional to the additional resistance, is, as he states, mainly based on the effect in copper, silver and gold. This work

is amplified in a later publication (Kapitza 1933). We shall note that these metals also behave differently in showing no appreciable α effect at appropriately low temperatures.

2—Meisener and Scheffers (1929, 1930)

The work of these authors on gold single crystals is thought not to have much bearing on ours, for the reason just noted.

3—De Haas and van Alphen (1933)

Their results for cadmium at 77.4 and at 20.4° K. and ours agree excellently. The curve at nitrogen temperature may be superposed on ours within experimental error. At 20.4° K. we derive from their curve of $\Delta R/R_0$ with H

$$H_k = 2.6 \text{ kilogauss, } 10^4 \times b = 12.7 \text{ kilogauss}^{-1}$$

which fit well on to our curves of figs. 4 and 5.

They also examined their cadmium specimen at solid nitrogen temperature (63.8° K.) and here we deduce from their results,

$$H_k = 7 \text{ kilogauss, } 10^4 \times b = 6.5 \text{ kilogauss}^{-1},$$

the resistance ratio being given as $R/R_0 = 1919 \times 10^{-4}$.

At the temperature of solid hydrogen we observe in their figures a slight curvature of the graph of ΔR with H , indicating the onset of the α effect.

4—De Haas and Blom (1934, 1935)

De Haas and Blom (1934, 1935) examined in great detail the behaviour of gallium single crystals. While principally they investigated the dependence on orientation, they published curves of $\Delta R/R_0$ with H at 77.4, 63.8, 56.5, 49.8, 20.4 and 14.2° K. for the two principal positions of their crystal, denoted by $\alpha = 0$ and $\alpha = 90^\circ$.

We see the characteristics most plainly in their curves at 49.8° K. For $\alpha = 90^\circ$ the curves show the behaviour found by Kapitza, with $H_k = 4$ kilogauss at this temperature. The $\alpha = 0$ curve shows a tendency to saturation. We would infer from this a rather low critical field (about 1 kilogauss). For higher temperatures, the critical field for $\alpha = 90^\circ$ is higher, and the downward curvature of the $\alpha = 0$ curve is less. For hydrogen temperatures, a parabolic curvature is to be seen for both orientations—we take this to be the α effect.

Including with these their measurements at 9.9, 4.2, and 1.35° K., we may plot a curve showing $\Delta R/R_0$ in a field of 22.3 kilogauss for the two orientations. We observe a point of inflexion in this case (cp. fig. 9) at 11° K.,

while we found this at 6.5° K. for cadmium. We observe that these temperatures are approximately in the ratio of the effective Debye temperatures for the two metals (for Cd 120° K. (de Haas, de Boer and van den Berg 1935), for Ga 210° K. (Dr Blom private communication)) on which the resistance also depends (Gruneisen 1927). As de Haas and Blom observe (1937), the value of the increase of resistance increases as R/R_0 decreases, becoming constant where the resistance becomes constant.

In their latest publication, de Haas and Blom (1937) also give curves of ΔR_{\parallel} with H for their gallium crystal. As the temperature is lowered, the saturation effect which we found becomes evident. At 10.4° K. (their lowest temperature) ΔR_{\parallel} is very much less than ΔR_{\perp} : the ratio is about 1:50, while for higher temperatures, e.g. 49.8° K., the parallel effect was actually greater than the perpendicular. This anomalous behaviour was also observed by Kapitza for gallium at nitrogen temperatures.

5—Other experiments relative to the α effect

The dependence just noted of the temperature of onset of the α effect on the Debye temperature θ is borne out in part by evidence which may be gleaned from other sources. We observe, for example, that de Haas and van Alphen's (1933) curve for zinc at 20.4° K. exhibits a slight parabolic curvature similar to that seen in their curve for cadmium at 14.2° K., but which is absent from their curves and ours for cadmium at 20.4° K. The effective θ value for zinc is not accurately known*, but it is probably higher than for cadmium, since zinc is of lower atomic number.

The case of aluminium, which de Haas and van Alphen also examined, is interesting on account of the high Debye temperature (390° K.). The resistance ratio and the change of resistance are both constant in the temperature range 20.4–14.2° K. We should expect the α effect to be fully established, but the curves tend to perfect linear asymptotes. We assume that the α effect is either absent, or much smaller, in Al than in Cd.

The curve for silver ($\theta = 230^\circ$ K.) obtained by van Alphen (1933) and given by de Haas and Blom (1937) suggests interestingly an approach to

* Mott and Jones, in their recent book "The Theory of the Properties of Metals and Alloys", quote for the characteristic temperature of zinc the range 200–300° K., and for cadmium the value 172° K. In view of the known dependence of the "effective Debye temperature" for non-cubic metals on the temperature of the experiment, we prefer the value previously quoted (120° K.) for cadmium, since this is given by de Haas, de Boer, and van den Berg (1935) as referring to the range (1–20° K.) in which we are interested. The values given by Mott and Jones suggest, however, that the effective θ value would be higher at a given temperature for zinc than for cadmium.

saturation, as would be expected from Frank's formula in the absence of an α effect.

At 20.4°K. , Kapitza (1933) found for copper a pure linear effect. θ for copper may be taken as 315°K. , so that the α effect should have been appreciable at hydrogen temperatures in these experiments.

Thus the evidence available suggests that these metals, which appeared to behave differently in regard to the dependence of the critical field on purity, also are different from the non-cubic metals in not showing an appreciable α effect.

These experiments have been carried out in the Royal Society Mond Laboratory at Cambridge. They were started at the suggestion of Professor Kapitza, and continued under the direction of Professor Lord Rutherford and Dr J. D. Cockcroft, to all of whom my best thanks are due for their continual interest and encouragement.

Among the personnel of the Laboratory, who have all from time to time made helpful suggestions in friendly discussion, I would like especially to thank Dr R. Peierls, whose theoretical knowledge, and Mr H. Pearson, whose technical advice, have been most valuable. Mr Pearson also prepared the liquefied gases used in the work, and Mr E. Laurmann kindly made the various spectroscopic analyses mentioned in the text. I am also very grateful to Dr J. W. Blom of Leiden for a valuable discussion, in the course of which he acquainted me with certain results before publication.

I am indebted to the Department of Scientific and Industrial Research for a grant which has made it possible for me to carry out this work. It gives me pleasure to record here my appreciation of their generous assistance.

SUMMARY

Apparatus is described which was used to measure the increase of electrical resistance of specimens of cadmium at low temperatures, in magnetic fields up to 26 kilogauss.

At temperatures above 20°K. , Kapitza's observation, that for strong fields a linear variation of resistance with field strength replaced the initial square law, was confirmed.

Experiments at 20.4°K. with polycrystalline wires of differing purity showed that the coefficient b of the linear law was the same for all specimens, and that the critical field from which the linear increase started was proportional to the specific resistance of the specimen. Experiments at 77°K.

supported these findings. A disagreement with Kapitza's conclusions is discussed.

At helium temperatures a new effect was found. In addition to the linear effect mentioned, a square-law increase was observed, producing a very large increase of resistance in pure specimens. The increase of resistance could be represented as $a \times H^2 + b \times H$, where b was again constant, and a was inversely proportional to the residual resistance of the specimen.

Experiments were made in which the temperature varied between 3 and 21° K. These showed that the coefficient a is constant at the lowest temperatures, but falls rapidly above 4° K. to negligible values. b , however, varies only slightly with the temperature.

The a effect is absent when the magnetic field is parallel to the current, and the curve then shows a saturation effect, at helium temperatures.

The evidence available from other workers is discussed. It would seem that the a effect has previously been observed, though not recognized, in other metals of the non-cubic groups: but that the effect is absent, or much smaller, in the cubic metals, copper, silver, gold, and aluminium.

REFERENCES

- de Haas and van Alphen 1933 *Commun. Phys. Lab. Univ. Leiden*, No. 225a.
de Haas and Blom 1937 *Rapp. VII Congr. Int. du Froid* (in the Press).
— — 1934 *Commun. Phys. Lab. Univ. Leiden*, Nos. 229b, 231b.
— — 1935 *Commun. Phys. Lab. Univ. Leiden*, No. 237d.
de Haas, de Boer and van den Berg 1935 *Commun. Phys. Lab. Univ. Leiden*, No. 236d.
Frank 1931 *Rev. Mod. Phys.* 3, 1.
Gruneisen 1927 *Z. Phys.* 46, 151.
Hill 1934 *J. Sci. Instrum.* 11, 281, 246.
Kapitza 1927 *Proc. Roy. Soc. A*, 115, 658.
— 1929 *Proc. Roy. Soc. A*, 123, 202.
— 1933 *Leipzig. Vortr.*
— 1934 *Proc. Roy. Soc. A*, 147, 189.
Kapitza and Milner 1937 *J. Sci. Instrum.* 14, 165.
Meissner and Scheffers 1929 *Phys. Z.* 30, 827.
— — 1930 *Phys. Z.* 31, 574.
Meissner and Voigt 1930 *Ann. Phys., Lpz.* 7, 761.
Milner 1937 *Proc. Camb. Phil. Soc.* 33, 145.
Peierls 1930 *Leipzig. Vortr.* p. 75.
— 1931 *Ann. Phys., Lpz.*, 10, 97.
Schubnikow, de Haas and Blom 1935 *Commun. Phys. Lab. Univ. Leiden*, No. 237b.
van Alphen 1933 *Thesis, Leiden*.
-

Theory of Electrical Breakdown in Ionic Crystals

By H. FROHLICH

Wills Physical Laboratory, University of Bristol

(Communicated by N. F. Mott, F.R.S.—Received 2 January 1937)

1—INTRODUCTION

If an electrical field higher than a certain critical strength F is applied to an ionic crystal, the insulation breaks down. If the temperature is above a certain critical value T_0 (usually of the order $100^\circ\text{C}.$), F decreases very rapidly with temperature, and the breakdown takes place some seconds after the application of the field. Wagner (cf. Semenoff and Walther 1928) has shown that in this case the breakdown is due to the Joule heat generated by ionic conduction, which causes local melting.

For temperatures less than T_0 , on the other hand, the breakdown takes place in a time of the order 10^{-8} sec. (Rogowsky 1928) and the variation of F with temperature is very much smaller than in the case of heat breakdown.* The phenomenon in this case is referred to as electrical breakdown, any melting of the crystal being ruled out by the short times involved.

To explain the electrical breakdown, various theories have been proposed. The mechanical theory (cf. Semenoff and Walther 1928) assumes that the breakdown is due to mechanical rupture of the crystal, caused by the forces which the electrical field exerts on the ions. According to this theory the electrical strength, like the mechanical strength, should depend very strongly on cracks and other crystal imperfections. Experiments by v. Hippel (1932) have shown, however, that the breakdown field is almost the same for different specimens and does not depend on their source or method of preparation.

On the other hand, Joffe has assumed (cf. Semenoff and Walther 1928) that the breakdown is due to an ionization of the ions by the moving ions, carrying the current. This theory does not allow, however, for the short time in which the breakdown occurs. v. Hippel and others (v. Hippel 1935) have therefore suggested the following mechanism: At any temperature a few electrons will be in the "conduction level", i.e. free to move through the lattice. If these can gain enough energy from the field to ionize the

* Exact measurements of F in a homogeneous field have been made only at room temperature (v. Hippel 1935).

(negative) ions of the lattice, the number of free electrons will increase very rapidly. This will then lead to the breakdown similarly as in gases. v. Hippel (1935) supports these ideas by showing that the breakdown always occurs in such directions in the lattice, along which an electron has to surmount the lowest potential walls. (For a NaCl lattice, for example, this is the case in the 110 direction.)

In order to calculate the energy gained by the electron from the field, we need to know the mean free path of the electron in the lattice. In this paper, therefore, we make a quantitative calculation of the mean free path, and hence the breakdown field, on the assumption that the phenomenon is due to electrons.*

2—THE CONDITION FOR BREAKDOWN

In an ionic crystal a very few electrons will always be in the conduction level. Their number increases with increasing field strengths but for small field strengths an equilibrium number is reached (cf. below). We shall assume these electrons to behave as though free (i.e. we neglect the effect of the lattice field). We denote the kinetic energy of such an electron by E , and assume that E is of the same order of magnitude but less than the energy \mathcal{F} required to excite or ionize the ions of the crystal (~ 5 e-volts). When such an electron is deflected by a collision with the lattice vibrations the gain or loss of energy is of the order $\hbar\nu$ (~ 0.02 e-volt), so that the collision is nearly elastic as in the theory of metals. Under these conditions the current density per electron is given by the usual formula

$$I = \frac{e^2 \tau F}{Vm}, \quad (1)$$

where F is the field strength, τ is the time of relaxation† and V is the volume of the crystal. Therefore, the energy A , transferred per second from the field to the electron, is given by

$$A = IFV = e^2 F^2 \tau / m. \quad (2)$$

The calculation of τ in § 4 gives

$$A = \text{const. } F^3 E^{\frac{1}{2}}.$$

* v. Hippel (1935) has already proposed some ideas for the interaction of an electron with an ionic crystal. The following calculations, however, lead to another conception of this interaction.

† τ is defined as the time in which the component of the momentum of the electron in a certain direction is reduced to the e th part (cf. § 4).

The electron also transfers a certain energy B per second to the lattice. In § 5 we shall find

$$B = \text{const. } E^{-1}.$$

The rate of loss of energy is thus

$$B - A = \text{const. } E^{-1} - \text{const. } F^2 E^{\frac{1}{2}}.$$

We see, therefore, that for low energies ($E < E'$) the electron loses energy and for high energies ($E > E'$) it gains energy. The critical energy E' , where* $A = B$ depends on the field strength F :

$$E' \propto 1/F.$$

Now the condition for breakdown is that no stationary state for the electron distribution can be reached, i.e. that an electron, with energy less than \mathcal{F} (the ionization potential), is capable of gaining energy from the field but that no reverse process exists. Hence the condition for breakdown is

$$E' \leq \mathcal{F}.$$

If this is not the case, the field cannot cause any instability because it cannot produce any *more* electrons capable of ionization. Thus, the critical field strength F where the breakdown begins can be calculated from the condition

$$E' = \mathcal{F},$$

$$\text{i.e.} \quad A(E, F) = B(E), \quad E = \mathcal{F}. \quad (3)$$

From the above considerations it follows that in a field, higher than the breakdown field F , the electron has in general to suffer quite a number of collisions before reaching the energy \mathcal{F} . We should notice, however, that even in a field weaker than the breakdown field it might happen that an electron having by chance an energy nearly equal to \mathcal{F} , and moving in the direction of the field, is accelerated in the field in such a way that it reaches between *two* collisions the energy \mathcal{F} . Processes of this sort do not lead, however, to any instability. They have only the effect that the number of electrons in the conductive levels increases with increasing field strength (but there exists a stationary state) and that therefore also the conductivity increases. For field strengths near to the breakdown field this increase of the conductivity has been observed. But as v. Hippel (1935) has shown, at the breakdown field the current density increases almost discontinuously. Thus the breakdown cannot be considered as a continuous increase of the ordinary current.

* An electron with the energy E' is in an unstable equilibrium.

3—INTERACTION BETWEEN THE LATTICE VIBRATIONS AND THE ELECTRON

The investigations of Born and v. Karman (Born and Göppert-Mayer 1933) on the elastic vibrations of a diatomic polar lattice show that the normal modes can be divided up into two branches, the acoustic and the optical branch. The number of normal modes in each branch is $3N$ (i.e. the total number is $6N$), where $2N$ is the number of ions.

Each elastic deformation of the lattice is connected with a certain polarization. Each elastic wave, therefore, corresponds to a certain "polarization wave". The oscillations of the optical branch are those in which two neighbouring ions of opposite sign vibrate in opposite directions. These waves correspond therefore to long polarization waves. On the other hand, in the acoustic branch, neighbouring ions of opposite sign have almost the same displacement. These waves correspond, therefore, to polarization wave-length, nearly equal to the lattice constant.†

According to Born and v. Karman (Born and Göppert-Mayer 1933) the oscillations of the optical branch have all nearly the same frequency ν (the Reststrahlen frequency). Since the optical branch gives much the greater polarization of the lattice, it gives a much stronger interaction with the electron. Thus it will involve only a small error if we give the Reststrahlen frequency ν to all waves.

For a long polarization wave (wave number w) the polarization P_w per unit volume may be obtained as follows: Let u_+ , u_- denote the displacements of the ions of the two signs in the neighbourhood of a given point, and let

$$u_w = u_+ + u_-.$$

Then clearly

$$P_w = eu_w/2a^3, \quad (4)$$

where $2a^3$ is the volume of the unit cell, i.e. a the distance between neighbouring ions.

For those waves in which the polarization varies very rapidly from point to point (short polarization waves, long elastic waves) we shall adopt the same device as that used by Debye in his theory of specific heats and treat the crystal as a continuous medium. We shall treat the elastic waves as giving rise to a polarization wave

$$u_w(x, y, z) = \frac{1}{(2N)^{1/2}} \{b_w e^{i(\mathbf{w}, \mathbf{r})} + b_w^* e^{-i(\mathbf{w}, \mathbf{r})}\}, \quad (4a)$$

where

$$b_w = \text{const. } e^{2\pi i w t}.$$

† The wave number of an elastic wave differs from that of a polarization wave by a vector in the reciprocal lattice.

Since for each wave number w there are two transverse waves and one longitudinal one, w will have all values between 0 and w_0 , given by†

$$6N = 3 \frac{2a^3 N}{(2\pi)^3} \frac{4\pi}{3} w_0^3,$$

$$\text{i.e.} \quad w_0 = \frac{(6\pi^2)^{1/3}}{a} \simeq 2^{1/3} \frac{\pi}{a}. \quad (5)$$

The value $w = 0$ corresponds to an infinite wave length, i.e. to a constant polarization.

For our further considerations it will be necessary to know the total energy of a polarization wave. Let M^+ and M^- denote the masses of the ions of the two signs. In our continuum theory, we shall have to assume the masses and charges of each sort of ion to be uniformly distributed. If there is no displacement the density of positive and negative charges is the same all over the continuum. Since the force acting on positive charges is opposite to that acting on negative charges, and since both u_+ and u_- have the same dependence on time, it follows that

$$M^+ u_+ = M^- u_-.$$

Using this relation, it is possible to calculate the total energy in the same way as has been done for the vibrations of a monatomic lattice.† One thus finds that the total energy is represented by the energy of an harmonic oscillator with a frequency ν , amplitude

$$X_w = b_w + b_w^*, \quad (6)$$

$$\text{momentum} \quad Y_w = -2\pi i \nu M (b_w - b_w^*), \quad (7)$$

$$\text{and mass } M, \text{ where} \quad \frac{1}{M} = \frac{1}{M^+} + \frac{1}{M^-}. \quad (8)$$

In order to consider the interaction of an electron with the dielectric, it is convenient to express the total energy in a Hamiltonian form. The Hamiltonian is composed of three terms

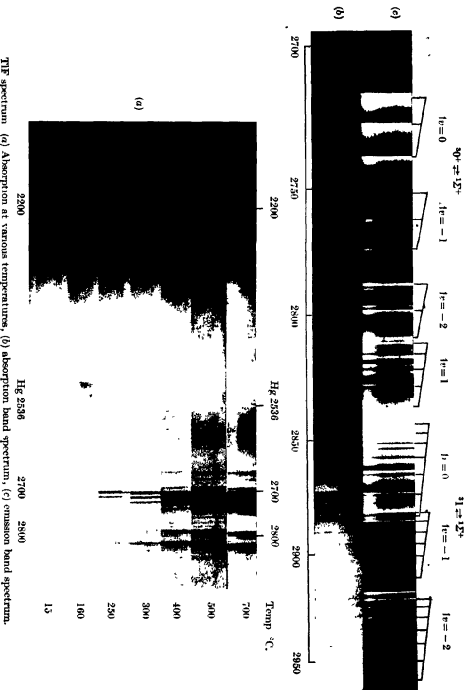
$$H = H_{\text{osc.}} + H_{\text{el.}} + W.$$

$H_{\text{osc.}}$ is the sum of the Hamiltonians of all the oscillators‡

$$H_{\text{osc.}} = \sum_w \left(\frac{Y_w^2}{2M} + 2\pi^2 \nu^2 M X_w^2 \right).$$

† (Cf. Peierls 1929.) In our case the kinetic energy is the sum of the kinetic energies of the positive and negative ions. The total energy is twice the kinetic energy as for any kind of harmonic oscillation.

‡ The interaction energy between the waves is zero, if they are assumed to be exactly harmonic (cf. Peierls 1929).



limited by the existence of the unstable level causing predissociation. In emission the distribution among the upper v' levels depends upon the particular mode of excitation but if the number of levels is limited by predissociation then the same levels must be occupied as those involved in absorption. Once this is granted then the equivalence of the transition probabilities in emission and absorption will ensure that the same v' levels are involved as in absorption. Thus the transitions are completely reversible and the emission band spectrum plate will be a positive of the absorption plate. It is probable that such a similarity in the intensity of emission and absorption spectra may be taken as an indication of predissociation.

The intensity distribution in the individual band systems is shown in Table IV. The Condon curve drawn through the strongest bands in each

TABLE IV—INTENSITY DISTRIBUTION

System $^30 \rightleftharpoons ^1\Sigma^+$					
$v' \backslash v''$	0	1	2	3	4
0	10	6	3		
1	3	10	7	8	
2	1	2	9	4	6
3		0			

System $^31 \rightleftharpoons ^1\Sigma^+$														
$v' \backslash v''$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	10	6	5											
1	7	10	6	5	4									
2				9	4	4P 2R		4						
3				5	8			4	3	2				
4					4	7	3		2	2	1			
5						2	5	2	2	2	1			
6								2		1	1	0		
7												1	0	0
8													1	0

Above line figures refer to P heads
Below line figures refer to R heads

system is narrow as is expected when $\omega'' \approx \omega'$. The rapid change in intensity of the P and R branches in system $^31 \rightleftharpoons ^1\Sigma^+$ can be followed in this table. This change is associated with the change-over of the direction of degradation of the heads and is another consequence of the nearness of the ω'' and ω' values and of the B'' and B' values. When B' is of the same order as B''

H_0 is the Hamiltonian of the electron, moving in the field of the lattice all ions are exactly in their equilibrium positions. Since we have assumed the electron to behave like a free electron, H_0 is given by

$$H_0 = p^2/2m,$$

where p is the momentum of the electron.

W is the interaction energy between the electron and the polarization waves. Thus

$$W = e\mathcal{E}\phi_w,$$

where ϕ_w is the potential of a single wave with wave number w at the position of the electron. ϕ_w is determined by

$$\nabla^2\phi_w = 4\pi\text{div}\mathbf{P}_w. \quad (9)$$

From this equation it follows that only longitudinal waves have an interaction with the electron. Using equation (4), we find for the solution of (9) according to (4a), (6) and (7)

$$\begin{aligned} \phi_w &= -\frac{4\pi e}{2\omega^2(2N)^{\frac{1}{2}}} \{b_w^* e^{i(\mathbf{w}, \mathbf{r})} - b_w^* e^{-i(\mathbf{w}, \mathbf{r})}\} \\ &= \frac{2\pi e}{2\omega^2(2N)^{\frac{1}{2}}} \left\{ X_w \sin(\mathbf{w}, \mathbf{r}) + \frac{2\pi\nu M}{X_w} \cos(\mathbf{w}, \mathbf{r}) \right\}. \end{aligned} \quad (10)$$

From the Hamiltonian H , we may, for instance, calculate the classical polarization P of the dielectric due to an electron. We find $P \propto e/r$ for distances $r \gg a$. For $r \rightarrow 0$, P tends to a certain finite value in contrast to the classical picture where e/r is considered a constant and $P \rightarrow \infty$ as $r \rightarrow 0$.

4—TIME OF RELAXATION

According to the quantum theory, the momenta p and X_w have to be replaced in the usual way by operators. The energy of each polarization wave consists then of a number of quanta $\hbar\nu$.

For the calculation of the time of relaxation, we may treat W as a perturbation. Then, the zero order approximation of the wave function of the total system is given by a product

$$\psi = \prod_w \chi(X_w, n_w) \psi(\mathbf{r}, \mathbf{k}),$$

and the total energy by

$$U = \sum_w (n_w + \frac{1}{2}) \hbar\nu + E_0.$$

ϵ is the dielectric constant.

the separation $\nu_H - \nu_O$ between the head and origin of a band which is given by the expression

$$\nu_H - \nu_O = \frac{-(B' + B'')^2}{4(B' - B'')}$$

becomes very large or even changes sign. When this occurs the band is either headless, or a reversal in the direction of degradation is found, as with the 3, 3; 2, 1; 2, 2; and 3, 5 bands. Assuming that for these levels

$B'' = B'$, an estimate of the ratio $\frac{\alpha'}{\alpha''}$ can be made where α is the vibrational

coefficient in the relation $B = B_e - \alpha v$; $\frac{\alpha'}{\alpha''}$ turns out to be about 2, a value

which is unusually high. This large difference in the α 's means that although B' and B'' may be almost equal for a particular band, they rapidly become widely different so that $\nu_H - \nu_O$ which depends upon $B' - B''$ rapidly alters too. This results in a sudden spacing out or crowding together of the lines of a branch from band to band with a sequence, with the consequent alteration in the branch intensities.

ISOTOPE EFFECT

According to Aston (1932) Tl has two isotopes of mass 205 and 203, their abundance ratio being 2.4 : 1 so that it should be possible to detect the weaker (203) by the band spectrum method, but no such discovery has yet been reported. As F has no known isotopes the spectrum of TlF is particularly suitable for the study of the Tl isotopes. The separation of the two isotopic heads of a given band ν' , ν'' is given by

$$\nu' - \nu = (\rho - 1) \left\{ (\nu - \nu_e) - x' \omega' \left(v' + \frac{1}{2} \right)^2 + x'' \omega'' \left(v'' + \frac{1}{2} \right)^2 \right\},$$

where ν' refers to the less abundant molecule and $\rho = \sqrt{\frac{\mu}{\mu'}} = 1.0004$ for TlF.

The maximum isotope separation calculated for the $^3\text{I} \leftarrow ^1\Sigma^+$ system is -1.2 cm^{-1} and refers to the 7, 12 band. Examination fails to reveal any sign of a weaker isotope head, but it may be concealed by the stronger since the Tl_{205}F head will be on the long-wave side of the Tl_{203}F head, which is itself shaded to the red. Violet-degraded bands would therefore be more suitable for the detection of the Tl_{203} isotope, but these are few and are so near the system origin that the isotope separations are too small to be observed.

conservation of energy and momentum, it follows† that the angle α between \mathbf{k} and \mathbf{w} is given by

$$\cos \alpha = -\frac{w}{2k} \pm \frac{2m}{\hbar^2} \frac{\hbar\nu}{2kw} \quad (+ \text{ for absorption, } - \text{ for emission}). \quad (13)$$

For most oscillators w , the second term in (13), is small.

We can now calculate the time of relaxation τ . Let the field F be in the x direction, so that the change of the x component k_x of the wave number of the electron by the field is

$$\left(\frac{\partial k_x}{\partial t}\right)_{\text{field}} = \frac{eF}{\hbar}.$$

Let $\Delta k_x(\mathbf{w})$ be the mean change of k_x by one collision with the oscillator \mathbf{w} . Then, τ is defined as (Fröhlich 1936, § 14)

$$\frac{k_x}{\tau} = -\sum_{\mathbf{w}} \Delta k_x(\mathbf{w})(\Phi_w^a + \Phi_w^e). \quad (14)$$

This definition leads immediately to equation (1), since the right-hand side is the mean change of k_x per second, due to collisions, i.e.

$$\frac{k_x}{\tau} = -\left(\frac{\partial k_x}{\partial t}\right)_{\text{collision}}.$$

Equation (1) then follows from the condition for a stationary state,‡

$$\left(\frac{\partial k_x}{\partial t}\right)_{\text{field}} + \left(\frac{\partial k_x}{\partial t}\right)_{\text{collision}} = 0.$$

Let us introduce polar co-ordinates (w, θ, ϕ) in \mathbf{w} space, with the axes parallel to \mathbf{k} , i.e. with $\theta = \alpha$. Since the Φ_w 's do not depend on ϕ , we may take the average of $\Delta k_x(\mathbf{w})$ in equation (14) over the azimuth ϕ . The result is (Fröhlich 1936, § 13)

$$\overline{\Delta k_x(\mathbf{w})} = -\frac{w^2}{2k^2} k_x.$$

Replacing the sum in (14) by an integral, we find for τ (cf. equation (5))

$$\frac{1}{\tau} = \frac{2a^3 N}{(2\pi)^3} \int_0^{2\pi} d\phi \int_{-\pi}^{\pi} \sin \theta d\theta \int_0^{w_0} \frac{w^2}{2k^2} (\Phi_w^a + \Phi_w^e) w^2 dw.$$

The integration over ϕ gives 2π . The simplest way to carry out the integration over θ is (cf. Bloch 1928, p. 589) to introduce ξ as variable instead of θ .

† Cf., for example, Fröhlich (1936, § 13), considering that all oscillators have the same energy.

‡ In fact, the state is not stationary, but quasi-stationary, because of the exchange of energy of the electron with lattice and field (cf. § 2).

Here, \mathbf{k} is the wave number and ψ the wave function of the electron

$$\psi = e^{i(\mathbf{k}, \mathbf{r})} / V^{\frac{1}{2}},$$

$E_{\mathbf{k}}$ is its energy:

$$E_{\mathbf{k}} = \hbar^2 k^2 / 2m,$$

n_w is the number of quanta $\hbar\nu$ of the oscillator w , and $\chi(X_w, n_w)$ is its wave function.

We now calculate the probability Φ_w per second that the electron makes a transition to the state \mathbf{k}' , coupled with a transition of the oscillator w from n_w to n'_w . By Dirac's perturbation theory, Φ_w is given by†

$$\Phi_w = \frac{1}{\hbar^2} |M_{\mathbf{k}\mathbf{k}'w}|^2 \frac{\partial \sin^2 \xi t}{\partial t \xi^2},$$

where

$$\hbar\xi = E_{\mathbf{k}'} - E_{\mathbf{k}} + (n'_w - n_w) \hbar\nu$$

and $M_{\mathbf{k}\mathbf{k}'w} = \frac{1}{V} \int \chi^*(X_w, n'_w) e^{-i(\mathbf{k}', \mathbf{r})} e\phi_w \chi(X_w, n_w) e^{i(\mathbf{k}, \mathbf{r})} dX_w d\mathbf{r}.$

Introducing ϕ_w from equation (10) and using the well-known matrix elements for an oscillator, we find that the electron may either absorb ($\propto \Phi_w^a$) or emit ($\propto \Phi_w^e$) a quantum $\hbar\nu$ (as in the theory of metals). The matrix elements can easily be worked out. In the case of emission, $M_{\mathbf{k}\mathbf{k}'w}$ is different from zero only if \mathbf{k}' is determined by

$$\mathbf{k}' = \mathbf{k} - \mathbf{w}. \quad (11)$$

The matrix element is then

$$M_{\mathbf{k}\mathbf{k}'w} = \frac{2\pi e^2}{a^2 w (2N)^{\frac{1}{2}}} \left[\frac{\hbar(1 + n_w)}{4\pi M\nu} \right]^{\frac{1}{2}}$$

and hence

$$\Phi_w^e = \left(\frac{2\pi e^2}{a^2 w} \right)^2 \frac{1 + n_w}{2M\hbar\nu} \frac{\partial \sin^2 \xi t}{\partial t \xi^2}. \quad (11a)$$

In the case of absorption, we find similarly

$$\mathbf{k}' = \mathbf{k} + \mathbf{w} \quad (12)$$

and

$$\Phi_w^a = \left(\frac{2\pi e^2}{a^2 w} \right)^2 \frac{n_w}{2M\hbar\nu} \frac{\partial \sin^2 \xi t}{\partial t \xi^2}. \quad (12a)$$

\mathbf{k}' , therefore, is completely determined by conservation of momentum [(11) and (12)]. Φ_w is big only if $\xi = 0$ (conservation of energy). From the

† By taking the average over a small energy interval, it can easily be seen that Φ_w is independent of the time.

As this integral diverges for $w = 0$, we must find the lower limit w' of w . This can be obtained from equation (13), using the fact that always

$$|\cos \alpha| \leq 1.$$

Since for small w the first term in equation (13) is small, w' is given by*

$$w' = \frac{2\pi m\nu}{\hbar k}.$$

$$\text{We then obtain } B = \frac{\hbar\nu}{\tau_0} \frac{4k^2}{w_0^2} \log \gamma = \frac{2^{\frac{1}{2}}\pi e^4 m^{\frac{1}{2}}}{M a^3 E^{\frac{1}{2}}} \log \gamma, \quad (17)$$

where

$$\gamma = \frac{w_0}{w'} = \frac{\hbar k}{2^{\frac{1}{2}} m \nu a} = \frac{(2E)^{\frac{1}{2}}}{2^{\frac{1}{2}} m^{\frac{1}{2}} \nu a}.$$

6—BREAKDOWN FIELD

We calculate now the breakdown field F from equation (3). Using equations (2), (15), (16), (17) and (5), we obtain

$$eF = \frac{2^{\frac{1}{2}}\pi}{4} \frac{e^4}{a^4 \mathcal{F} M} \left(\frac{m\hbar \log \gamma}{\nu} \right)^{\frac{1}{2}} \left(1 + \frac{2}{e^{\hbar\nu/kT} - 1} \right)^{\frac{1}{2}}. \quad (18)$$

As τ is proportional to the mean free path l (l_0 = mean free path for $T = 0$), we notice that

$$eF \propto \frac{1}{(l_0)^{\frac{1}{2}}}. \quad (19)$$

7—DISCUSSION

Equation (18) contains no arbitrary constants so that we may calculate theoretically the absolute value of F . The energy \mathcal{F} will be taken from the first maximum of the ultra-violet absorption band. For the alkali halides, the latter has been measured by Hilsch and Pohl (1930). If λ is the wave-length of this first maximum, \mathcal{F} is given by

$$\mathcal{F} = \hbar c / \lambda.$$

We shall insert this expression into equation (18). Finally we shall express F in V/cm., λ and a in Angströms, $\hbar\nu$ in 10^{-2} V and M (cf. equation (8)) in units of the hydrogen mass. We then obtain

$$F = 1.6(\log \gamma)^{\frac{1}{2}} \times 10^5 \frac{\lambda}{M a^4 (\hbar\nu)^{\frac{1}{2}}} \left(1 + \frac{2}{e^{\hbar\nu/kT} - 1} \right)^{\frac{1}{2}} \text{ V/cm.}$$

* As $w' \ll w_0$, it was correct to put $w' = 0$ in the integral for $1/\tau$.

Using the fact that n_w is the same for all oscillators, namely,

$$n_w = \frac{1}{e^{h\nu/kT} - 1},$$

we finally obtain*
$$\frac{1}{\tau} = \frac{1}{\tau_0} \left(1 + \frac{2}{e^{h\nu/kT} - 1} \right), \quad (15)$$

where
$$\frac{1}{\tau_0} = \frac{2^{\frac{1}{2}}\pi^2}{8\sqrt{2}} \frac{e^4\hbar}{m^{\frac{1}{2}}Ma^3\nu E^{\frac{1}{2}}}, \quad \text{if } E > \frac{2^{\frac{1}{2}}\pi^2}{8} \frac{\hbar^2}{ma^3}, \quad (16)$$

and
$$\frac{1}{\tau_0} = \frac{e^4 m^{\frac{1}{2}}}{2^{\frac{1}{2}}\hbar Ma^3\nu E^{\frac{1}{2}}}, \quad \text{if } E < \frac{2^{\frac{1}{2}}\pi^2}{8} \frac{\hbar^2}{ma^3}. \quad (16a)$$

The time of relaxation τ_0 at the absolute zero of temperature is entirely due to the zero-point oscillations. In Bloch's electron theory of metals (Bloch 1928), owing to the Pauli principle, an electron cannot loose energy, and so there are no collisions of the conductive electrons with the zero point oscillations. This is not the case here. An electron, with an energy higher than the energy of the conduction electrons, suffers collisions with the lattice vibrations† even at $T = 0$.

5—TRANSFER OF ENERGY TO THE LATTICE

At each collision, the energy $h\nu$ is transferred to the lattice with a probability Φ_w^e , or absorbed by the electron with a probability Φ_w^a . The energy B , transferred per second to the lattice, therefore, is given by

$$B = h\nu \sum_w (\Phi_w^e - \Phi_w^a).$$

Since (cf. equations (11a) and (12a))

$$\Phi_w^e/\Phi_w^a = (1 + n_w)/n_w,$$

B is independent of the temperature.

B can be evaluated by a method similar to that used in the calculation of τ . We thus find

$$B = \frac{h\nu}{\tau_0} \frac{4k^2}{w_0^3} \int \frac{dw}{w}.$$

* For $2k < w_0$, w_0 has to be replaced by $2k$ because of momentum and energy theorem. This leads to equation (16a).

† If, however, $E < h\nu$, the electron cannot emit the energy $h\nu$, and (15) has to be replaced by

$$\frac{1}{\tau} = \frac{1}{\tau_0} \frac{1}{e^{h\nu/kT} - 1}.$$

In this case, therefore, $1/\tau$ vanishes for $T = 0$.

increase of temperature, there are two other means of decreasing the mean free path l . The first is to introduce foreign ions into the lattice,* the second is to take very thin layers of thickness smaller than l . An increase of F in the first case has been found by v. Hippel (1935). Experiments on thin layers have been carried out by Joffe and co-workers (Joffe and Alexandrow 1933). They did not find an increase in F for thicknesses down to 0.7×10^{-4} cm. The mean free path, however, is only of the order 10^{-5} – 10^{-6} cm., so that an increase of F is not yet to be expected.

I should like to express my thanks to Professors Tyndall and Mott for their kind hospitality in their laboratory, and to Professor Mott for help in the preparation of the manuscript.

SUMMARY

The time of relaxation of an electron in an ionic lattice has been calculated. Hence the critical field F for electrical breakdown has been calculated quantitatively. Satisfactory agreement with the experiments of v. Hippel is obtained.

F increases: (1) with increasing temperature (for high temperature $F \propto \sqrt{T}$, if $T < T_0$, T_0 is the temperature where heat breakdown begins); (2) if foreign ions are introduced into the lattice (experiments by v. Hippel); (3) for layers of about 10^{-6} cm. thickness.

REFERENCES

- Bloch 1928 *Z. Phys.* 52, 555.
 Born and Göppert-Mayer 1933 *Handb. d. Physik*, 2nd ed. 24/2, 623.
 Fröhlich 1936 "Elektronentheorie der Metalle." Berlin.
 Hilsch and Pohl 1930 *Z. Phys.* 59, 812.
 v. Hippel 1932 *Z. Phys.* 75, 145.
 — 1935 *Ergebn. exakt. Naturwiss.* 14, 79.
 Joffe and Alexandrow 1933 *Phys. Z. Sowjet.* 2, 527.
 Nordheim 1931 *Ann. Phys., Lpz.*, 9, 607.
 Peierls 1929 *Ann. Phys., Lpz.*, 3, 1055.
 Rogowsky 1928 "Sommerfeld-Festschrift (Probleme der modernen Physik)." Leipzig.
 Semenoff and Walther 1928 "Die physikalischen Grundlagen der elektrischen Festigkeitslehre." Berlin.

* As in the theory of conductivity, the presence of foreign ions always diminishes l (Nordheim 1931) and not only in the case where the ionic radius of the foreign ion is smaller than that of the original ion, as it was assumed by v. Hippel.

Here, λ is the wave-length of the ultra-violet absorption, ν is the Reststrahlen frequency, a is the distance between two neighbouring ions of opposite sign, and M is given by equation (8) in terms of the atomic weight M^+ and M^- of the ions. For $E = \mathcal{F}$, $(\log \gamma)^{\frac{1}{2}}$ has nearly the same value for all alkali halides, namely, 2.6.

We shall now compare for the alkali halides the theoretical value of F at 300° K. with the experiments of v. Hippel (1935) carried out at room temperature. v. Hippel's values differ considerably from former measurements. For NaCl, for example, his value is 15×10^5 V/cm., whereas Semenoff and Walther (1928) give 5×10^5 V/cm. As mentioned in the introduction, v. Hippel has shown that the breakdown occurs always in the 110 direction whatever direction the field may have and that, therefore, the projection of the breakdown field on the 110 direction is always the same. In our model which does not account for any anisotropic effects, the breakdown occurs in the direction of the field. Our formula for F , therefore, corresponds to the experimental value of the field in the 110 direction. To obtain the value in the 100 direction (F_{100}), for which experimental values usually are given, we have to multiply F by $\sqrt{2}$, as already mentioned by v. Hippel: thus $F_{100} = F\sqrt{2}$.

TABLE I— F_{100} IN 10^5 V/CM.

	NaCl	NaBr	NaI	KCl	KBr	KI	RbCl	RbBr	RbI
Theory:									
$T = 0^\circ$ K.	6.9	6.1	4.9	3.8	3.0	2.5	2.7	2.0	1.4
$T = 300^\circ$ K.	10.7	10.6	9.3	6.6	5.6	5.1	4.8	4.2	3.1
Experimental:									
$T \propto 300^\circ$ K.	15	10	8	8	7	6	7	6	5
m^*/m	2.0	0.9	0.7	1.5	1.6	1.4	2.1	2.0	2.6

Considering the simplifications made in the theory, the agreement is satisfactory. The difference between theoretical and experimental values may be due to the "effective mass" m^* for an electron in the lattice field being generally greater than the mass of a free electron.

An important consequence of the theory is the dependence of F on temperature. For sufficiently high temperatures ($kT > h\nu$), F should be proportional to \sqrt{T} , provided that the temperature T_0 , where the heat breakdown begins, is not yet reached. (For $T > T_0$, cf. § 1.) As we have already stated, experiments in a homogeneous field have at present been carried out at room temperature only.

The increase of F with temperature is due to the decrease of the mean free path with temperature (cf. equations (19) and (15)). Apart from an

depends upon the existence of a hitherto unknown F level 5280 cm.^{-1} above the ground state.

The present paper describes both the absorption and emission spectra of TlF, and analyses of all the bands mentioned by Boizova and Butkow, together with others not found by them, are presented here. A completely different interpretation of the results is put forward.

EXPERIMENTAL

The absorption tube used was a 30 cm. length of pyrex tubing 1 cm. in diameter, open at both ends so as to avoid using windows (because of the destructive action of F and HF). This was surrounded by a Nichrome wire heating coil lagged with asbestos. A little thallium fluoride placed in the centre of the tube lasted 4–5 hr. before either distilling out or combining chemically with the glass. A hydrogen discharge lamp running at 0.1 amp. 2500 V served as a source of continuum. A similar length of tubing with a quartz window cemented on one end was used as an emission tube. This was connected at the other end to a Cenco Hyvac oil pump via a liquid air trap. External iron electrodes wound round the tube about 2 in. apart supplied the excitation energy from a $\frac{1}{2}$ kW oscillator working at a frequency of 10^7 cycles/sec. The fluoride had to be heated before the bands were developed in the high frequency discharge, the most prominent feature of which was the intense green colour due to the Tl line 5350.

The spectra were photographed on the Hilger E1 spectrograph which has an average dispersion of 5 Å/mm. in the region studied, whilst the continua were studied with a Hilger Small Quartz spectrograph. Ilford "Monarch", "Double Express", and Schumann plates were used in photographing the spectrum which was surveyed initially from 1900 to 8000 Å. Iron arc lines served as wave-length standards.

DESCRIPTION OF THE SPECTRUM

Absorption spectra at various temperatures up to 700°C. are reproduced in Plate 1. Absorption first takes place at a temperature around 160° when the bands at 2700 and 2800 appear together with a double-headed set at 2200 which has a continuum on its short wave side. This 2200 band system is not recorded by Boizova and Butkow. As the temperature increases, this 2200 continuum extends in both directions and ultimately joins on its short-wave side a far ultra-violet continuum (not shown in the plate), also not recorded by Boizova and Butkow. At about 500° there

The Spectrum of Thallium Fluoride

By H. G. HOWELL, PH.D., *University College, Southampton*

(Communicated by W. E. Curtis, F.R.S.—Received 7 January 1937)

[Plate 1]

INTRODUCTION

This study of the thallium fluoride spectrum was undertaken as part of a detailed investigation into the molecular spectra of the series of heavy diatomic fluorides HgF , TlF , PbF and BiF . Whereas the spectra of PbF (Rochester 1936) and BiF (Howell 1936), of which analyses have already been published, contain no very unusual features the TlF spectrum is particularly rich in them, so that it has seemed desirable to extend the original investigation in order to include the other halides of thallium.

The absorption spectrum of the fluoride has already been examined by Boizova and Butkow (1936), their findings being summarized below:

1—A continuum at 2200 Å appears when the absorption tube is at a temperature of 155° C. Its long-wave edge moves towards the red with increase of temperature, being at 2700 for the unsaturated vapour and at 3400 for the saturated vapour when the temperature is 280° C. They attributed this continuum to the dissociation of Tl_2F_2 . $\text{Tl}_2\text{F}_2 \rightarrow 2\text{TlF} + \text{kinetic energy}$.

2—Three continua appear at higher temperatures, their red edges being at 2680, 2516 and 2240.

3—A large number of bands degraded to the red are found around 2680. Some of these were fitted into a quantum scheme.

4—Further bands are found around 2800 some of which are degraded to the red others to the violet whilst a few shew both directions of degradation. No attempt was made to analyse these but they state that a frequency difference of 420 cm^{-1} occurs several times in this spectrum.

Apart from the 2200 continuum all these features are attributed to the TlF molecule. The three continua are interpreted as being associated with the dissociation processes:

1. $\text{TlF} + h\nu_1 = \text{Tl} + \text{F} + \text{kinetic energy}$,
2. $\text{TlF} + h\nu_2 = \text{Tl} (\text{excited}) + \text{F} + \text{kinetic energy}$,
3. $\text{TlF} + h\nu_3 = \text{Tl} + \text{F} (\text{excited}) + \text{kinetic energy}$,

but Boizova and Butkow admit that the correctness of the third process

The analysis of the bands around 2850 proved to be more difficult. Pairs of violet-degraded bands at 2845, 2880 and 2920 obviously formed two progressions with a 470 cm^{-1} difference and red-degraded bands with a sequence difference of approximately 80 cm^{-1} could be assigned to a different quantum scheme, but the first differences (ΔG) were very inconsistent. Little progress was made until measurements of the so-called "absorption" lines (on the emission plates) were substituted for those of the band heads. It was then found that the two violet-degraded bands at 2843 and some of the

TABLE I—SYSTEM $^3O^+ \rightarrow ^1\Sigma^+$

		<i>Vibrational analysis</i>									
v''	v'	0	1		2		3		4		
0	R	36839.7	461.4	36378.3	461.6	35916.7					
	Q	36809.8	470.2	36339.6	467.8	35871.8					
	RR	311.9		309.9		307.8					
	QQ	334.3		335.9		335.1					
1	R	37151.6	463.4	36688.2	463.7	36224.5	457.1	35767.4			
	Q	37144.1	470.8	36675.5	468.5	36207.0	463.3	35743.7			
	RR	307.5		303.9		299.7		294.2			
	QQ			311.9		311.2		310.5			
2	R	37459.1	467.0	36992.1	467.9	36524.2	462.4	36061.8	458.6	35603.2	
	Q			36987.4	469.2	36518.2	464.4	36054.2	459.5	35594.7	
	RR			284.2							
3	R			37276.1							

Separation of R and Q heads

v'	v''	0	1	2	3	4
0		29.9	38.7	44.9		
1		7.5	12.7	17.5	23.7	
2			4.7	6.0	7.6	8.5

"absorption" lines formed a sequence and that other similar sequences began at 2813, 2880, and 2920. These "absorption" lines were therefore considered to be the gaps produced by the missing line or lines around the band origins. However, the breadth of these gaps indicates that an unusually large number of lines, perhaps from 6 to 10, are of negligible intensity near the origin. For this reason, whereas settings could be made upon band heads to 0.1 \AA (or 1 to 1.5 cm^{-1}), band origins could not be determined to anything better than 0.3 \AA (3.5 cm^{-1}). However, in spite of this loss in accuracy in setting, the origin data are considered to be more significant than those obtained from band head measurements.

appears another continuum with a red edge at 2630. There is no sign at all of the 2516 continuum mentioned by Boizova and Butkow. It is suggested here that this absorption may have been due to some compound formed on the windows of their absorption tube as they remark that it was difficult to prevent them becoming opaque.

The plate also shows higher dispersion spectrograms of both emission and absorption spectra between 2700 and 2960 Å. It will be seen that the absorption is to a large extent a reversal of the emission. The most prominent features of the spectrum are the double-headed bands around 2710 and the complex band structure which stretches from 2800 to 2880. The bands around 2700 belong to the system analysed by Boizova and Butkow and it will be shown that their analysis of this system is incorrect. The other region contains red-degraded bands chiefly, but there are others, notably the strong pair at 2840, which are degraded to the violet. Among the bands in this region are what appear to be patches of continua broken by absorption lines on the emission plate and transmission lines on the absorption plate. The emission and absorption spectra are different in that the continua and the 2200 band system appear only in the latter.

THE BAND SYSTEMS

The bands around 2700 which are obviously to be classified as one system were analysed by Boizova and Butkow according to the formula

$$\nu = 36836 - (311\nu'' - 1.0\nu''^2) + (173\nu' - 10.6\nu'^2),$$

but a different analysis is submitted here. The double heads could not be accounted for by Boizova and Butkow but they are here considered to be *R* and *Q* heads. Using the *Q* head data good first differences are obtained in the quantum scheme shown in Table I which also gives *R* head data. This allocation of the double heads is supported by the variation of the *R* and *Q* head separations with ν , also given in Table I. This separation increases as ν'' increases and decreases as ν' increases in accordance with theoretical expectations. Another fact which gives support to the analysis proposed here is that all the bands in this region have been fitted into the quantum scheme.

The *Q* heads can be represented by the following formula:

$$\nu = 36869.5 + 360.65(\nu' + \frac{1}{2}) - 12.25(\nu' + \frac{1}{2})^2 - 475.00(\nu'' + \frac{1}{2}) + 1.89(\nu'' + \frac{1}{2})^2.$$

The accuracy of the fit of this formula to the observed values may be judged from the O-C column in Table V. The residuals range from -1.8 to 1.1 and have an algebraic sum of 0.4.

The complete vibrational matrix for this system is given in Table II. This contains *P* and *R* head and also origin data wherever the head or origin has been measured. An examination of the bands in the $\Delta v = 0$ sequence shows that the branch intensities vary widely. Thus, whilst the 0, 0 and 1, 1 bands have a strong *P* head and no other, the 2, 2 has a strong extended *R* branch which does not form a head, a conspicuous origin and a very short *P* branch which is also headless. The 3, 3 band has an extended *R* branch which now forms a head and a *P* branch which is headless but now not so short. The 4, 4 band has a strong and short *R* branch with a head and a weaker and extensive *P* branch, whilst the 5, 5 has a line-like *R* head and a long *P* branch. This changing over of the direction of degradation is thus accompanied by rapid fluctuations in branch intensities. This phenomenon is repeated in all the other sequences.

It can be seen from the vibrational scheme that

(a) Although the ΔG values using origin data are few, yet they vary regularly, unlike the ΔG values based upon *P* or *R* head data.

(b) The $\Delta G''$ values are of the same order as the $\Delta G''$ values of the previous system (using *Q* head data). This indicates that the lower level of both systems is the same and also justifies the identification of the "absorption" lines as band origins.

(c) The $\Delta G'$ values for the band origins are 405.0, 388.1, 371.9, and 353.9, whilst the first member of this series is missing as no band origins have been detected for the first v'' progression. However, as the second differences ($\Delta_2 G'$) are 16.9, 16.2, and 18.1 respectively for these cases, it is reasonable to assume in the first place that the first $\Delta_2 G'$ value will be 17.0 and that therefore the first $\Delta G'$ value will be 422.0. From this, the wave numbers of the origins of the bands of the first v'' progression have been calculated and thus the separation of *P* head and origin determined. These $\Delta G'$ values have been used to determine the following formula for the origins:

$$\nu = 35180.7 + 439.87(v' + \frac{1}{2}) - 8.60(v' + \frac{1}{2})^2 - 475.00(v'' + \frac{1}{2}) + 1.89(v'' + \frac{1}{2})^2.$$

The O-C values are given in Table V. They range from 2.5 to -2.9 and have an algebraic sum of -0.2. That the agreement between observed and calculated values is not so good as with the other system is attributed to the difficulty, already mentioned, of setting accurately on the band origins. The variation of $\nu_{\text{Head}} - \nu_{\text{Origin}}$ with v is also shown and is submitted as additional support for the analysis proposed here. It is found that with *P* heads the difference $\nu_H - \nu_O$ decreases as v'' increases and increases as v' increases whereas with *R* heads the reverse occurs, these results agreeing with theoretical expectations.

TABLE II.—SYSTEM $\lambda_1 = 12^\circ$, VIBRATIONAL ANALYSES

ν'	ν''	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	P	3537.5	449.4	3449.7	447.1	34320.4									
	PP	411.5	414.6	418.7											
1	P	3569.4	464.5	35109.2	459.4	34498.5	457.6	34181.7	455.5	33739.2					
	O	3583.8	472.8	35111.0	459.4	34449.7									
	FP		384.3	384.4	384.4	384.5	384.5	384.5	384.5	384.5					
	OO		459.0	459.0	459.0	459.0	459.0	459.0	459.0	459.0					
2	P			35039.1	454.8	34573.5	452.9	34139.0	451.3	33693.0					
	O			35119.5	454.8	34653.9	452.9	34219.4	451.3	33763.4					
	R			35018.0	449.1	33946.9	448.4	33511.3	446.8	33065.7					
	PP														
	OO														
3	P				367.8	368.4	368.1	368.4	368.1	368.4					
	O				3643.7	445.0	34572.7	437.7	34112.0	435.4	33657.8	433.6	33202.3		
	R				35410.2	443.8	34917.7								
	OO														
	RR														
4	O				342.0										
	R				33245.9	439.3	34836.6	434.0	34422.0	431.0	33977.0				
	OO				33581.2	438.0	34908.8	430.0	34480.2	437.6	34035.3				
	RR														
5	O														
	R														
	RR														
6	R														
	RR														
7	R														
	RR														
8	R														

Separation of R or P bands and origin

ν'	0	1	2	3	4	5	6	7
0	P (1.7)	P (6.0)	P (14)					
1	P (15)	P (9.7)	P (5.8)					
2	P (21.8)	P (19.8)	P (4.0)					
3	R (31.5)	R (44.0)	R (14.0)					
4	R (31.5)	R (44.0)	R (14.0)					
5	R (15.3)	R (11.6)	R (1.6)					
6	R (7.7)	R (15.4)	R (18.6)					

Bracketed values calculated.

As the separation of the origins of these two systems is only 1700 cm.^{-1} approximately, it is very probable that the upper levels form an electronic doublet. The third band system, at 2200 , consists of two v'' progressions only and these are shown in Table III. As it is so small no formula has been calculated for it, although it can be seen that ω_e for the upper state is of the

TABLE III—VIBRATIONAL ANALYSIS OF SYSTEM $^1\Pi \rightarrow ^1\Sigma^+$

v''	0	1	2	3	4	5	6
0	45481.2	470.2 45011.0	467.8 44543.2	464.8 44078.4			
		348.0	346.2	343.8			
1		45359.0	469.6 44889.4	467.2 44422.2	461.2 43961.4	460.4 43501.4	457.4 43044.4 456.4

order of 360 cm.^{-1} and also that the initial level (as the system is only obtained in absorption) is the ground level. Over-exposed emission plates revealed no trace of this system.

ELECTRONIC TRANSITIONS IN THALLIUM FLUORIDE

(a)—Stable Energy States

The spectra of the halides of the related group 3 elements In and Ga have been investigated by Miescher and Wehrli (1934*a, b*) and a comparison of their results with those obtained here shows that a strong resemblance exists between them all. Among the features common to all there may be mentioned

- (i) A doublet system in the ultra-violet.
- (ii) A singlet system to the short wave side of (i).
- (iii) Continua
- (iv) ω_e' and ω_e'' in a given system are of the same order.

Furthermore, although Miescher and Wehrli (1934*a, b*) make no mention of them, band origins similar to those of TlF are plainly visible in their reproductions. Now Holst (1934) has shown that the ground state of the Al halide molecules is $^1\Sigma$ and in all probability the halides discussed here have the same configuration. Miescher and Wehrli consider that the doublet systems in the Ga and In halide spectra are due to forbidden transitions between the ground level and certain components of a triplet level. If the ground state has an electron configuration $\sigma^2\pi^{41}\Sigma^+$ then excitation of one π electron will produce $\sigma^2\pi^3\sigma$ and states $^1\Pi$, $^3\Pi_0^+$, $^3\Pi_0^-$, $^3\Pi_1$, and $^3\Pi_2$ (inverted). When forbidden transitions occur as here, with relatively high intensity, it is probable that the coupling in the excited states resembles that of Hund's case *c* in which L and S are strongly coupled together to give a resultant J

The abrupt cut-off of the bands in the 2200 system is very marked. The upper ${}^1\Pi$ state has only two active levels owing to the predissociation caused by the presence of another unstable level B . Transitions from ${}^1\Sigma^+$ to this level B produces the strong continuum which ultimately overlies the band system. This is attributed by Boizova and Butkow to the dissociation of Tl_2F_2 , but against this interpretation is the vapour pressure evidence that the fluoride is diatomic in the vapour state.

To account for the far ultra-violet absorption it is necessary to postulate the existence of another unstable level C . It is possible that if the level A is not ${}^30^-$ it may be ${}^3\Sigma^+$ which is one of the states that can arise from the configuration $\sigma^2\pi^3\pi$ the others being ${}^1\Sigma^+$, ${}^1\Sigma^-$, ${}^1\Delta$, ${}^3\Sigma^-$ and ${}^3\Delta$. This would mean that the last π electron would have approximately the same binding energy as the σ in the former configuration, whereas it is usual for the level to be lower in energy value the smaller the l value, i.e. the levels arising from $\pi^2\pi$ should be higher than those arising from $\pi^3\sigma$. If the binding of these two electrons is the same then the repulsive level B can be identified as the level ${}^1\Sigma^+$ because if ${}^3\Sigma^+$ and ${}^30^+$, 31 have the same energy then it is not improbable that the separations ${}^3\Sigma^+ - {}^1\Sigma^+$ and ${}^30^- - {}^1\Pi$ are the same. That ${}^3\Sigma^+$ is the unstable level A and not another ${}^3\Pi$ (or ${}^3\Delta$) would also account for the occurrence of only one continuum and not two which would be expected from ${}^3\Pi_{0,1} \leftarrow {}^1\Sigma^+$. In view of the uncertainty, however, the designations A and B are retained for these levels in the energy diagram for this molecule, shown in fig 1. The dissociation products are not indicated as a study of the dissociation of this and the other halide molecules of Tl is not yet completed.

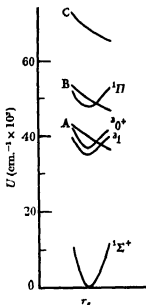


FIG. 1—Energy levels of TlF.

INTENSITY DISTRIBUTION

The distribution of intensity among the bands is the same in both emission and absorption. This is unusual as the factors which determine the intensity are different in the two cases. The existence of predissociation in the excited state is probably responsible for the phenomenon. As $\omega_e^* \approx \omega_e'$ the equilibrium r_e values will be of the same order so that transitions in absorption from the lower v'' levels will involve the same low values of v' —values which are

which has a component Ω along the internuclear axis, similar to an atom in a weak field. Here Λ loses its meaning and only Ω remains a rigorous quantum number. The permitted transitions are governed by the rules $\Delta\Omega = 0$ or ± 1 , and $+\leftrightarrow +$ $-\leftrightarrow -$ only. Thus the transitions ${}^3\Pi_0^- \leftarrow {}^1\Sigma^+$ and ${}^3\Pi_2^- \leftarrow {}^1\Sigma^+$ will not be allowed, leaving only ${}^3\Pi_0^+ \leftarrow {}^1\Sigma^+$ and ${}^3\Pi_1^- \leftarrow {}^1\Sigma^+$ —two systems as observed.

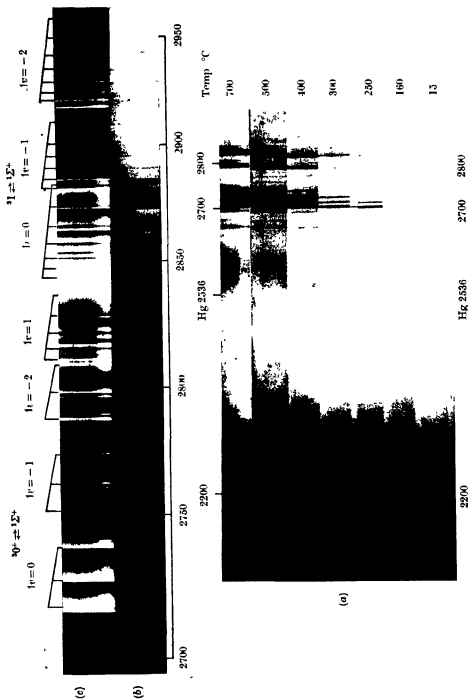
In TlF then, the farther ultra-violet system can be identified as the ${}^1\Pi \leftarrow {}^1\Sigma^+$ and the doublet system as the ${}^3\Pi_1^- \leftarrow {}^1\Sigma^+$ and ${}^3\Pi_0^- \leftarrow {}^1\Sigma^+$. With the Ga and In halides the ${}^1\Pi \leftarrow {}^1\Sigma$ system is reported to be stronger than the ${}^3\Pi_{0,1}^- \leftarrow {}^1\Sigma$ and this is in keeping with the assignment of transitions, as the forbidden transition should be the weaker. However, with TlF all the systems are of equal intensity so it is considered that the binding from Ga \rightarrow In \rightarrow Tl approaches more and more to Hund's case c as the molecule gets heavier. Consequently the Λ notation should be dropped and the levels denoted by their Ω values instead, thus ${}^3\Omega^+$, ${}^3\Omega^- \leftarrow {}^1\Sigma^+$. That the comparatively heavy Tl halides should have a coupling nearer to case c than the others is to be expected, for heavy atoms have a strong coupling between L and S .

(b)—Unstable Energy States

It is considered significant that no bands involving levels higher than $v' = 8$ for the ${}^3\Omega^-$ level and $v' = 3$ for the ${}^3\Omega^+$ level are observed. Transitions from $v'' = 0$ to these upper levels involve energy changes of 38,300 and 37,750 cm^{-1} respectively. Now the 2630 continuum has its red edge at approximately 38,050 cm^{-1} and therefore the repulsive level, here termed A , which is the upper level of this continuum, must cross both ${}^3\Omega^+$ and ${}^3\Omega^-$ levels and it is extremely probable that it produces predissociation in both. Something about the nature of this level A may be determined by examining the selection rules for predissociation. To produce predissociation two states must satisfy the two conditions

- (a) They must have the same multiplicity.
- (b) $\Delta\Omega = 0$ or ± 1 .

Consequently A must be a triplet level. It is tempting to assume that A is a third member of the triplet level to which ${}^3\Omega^-$ and ${}^3\Omega^+$ belong. In view of the inverted nature of this multiplet and the occurrence of the continuum on the short wave side of the other systems A would have to be ${}^3\Omega^-$. A transition ${}^3\Omega^- \leftarrow {}^1\Sigma^+$ involves a violation of the selection rule $+\leftrightarrow +$ only, but it can be argued that the rule merely stipulates a low probability of such a transition and it may be that the low intensity of this continuum in comparison with the band systems is some justification of this view.



Tl I spectrum (a) Absorption at various temperatures; (b) absorption band spectrum, (c) emission band spectrum.

limited by the existence of the unstable level causing predissociation. In emission the distribution among the upper v' levels depends upon the particular mode of excitation but if the number of levels is limited by predissociation then the same levels must be occupied as those involved in absorption. Once this is granted then the equivalence of the transition probabilities in emission and absorption will ensure that the same v' levels are involved as in absorption. Thus the transitions are completely reversible and the emission band spectrum plate will be a positive of the absorption plate. It is probable that such a similarity in the intensity of emission and absorption spectra may be taken as an indication of predissociation.

The intensity distribution in the individual band systems is shown in Table IV. The Condon curve drawn through the strongest bands in each

TABLE IV—INTENSITY DISTRIBUTION

$System\ ^30 \rightleftharpoons ^1\Sigma^+$

v'	v	0	1	2	3	4
0	0	10	6	3		
1	1	3	10	7	8	
2	2	1	2	9	4	6
3	3		0			

$System\ ^31 \rightleftharpoons ^1\Sigma^+$

v'	v	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	10	6	5											
1	1	7	10	6	5	4									
2	2	<div style="border-top: 2px solid black; border-left: 2px solid black; border-right: 2px solid black; padding: 2px; display: inline-block;"> 9 </div>			4	<div style="border-top: 2px solid black; border-left: 2px solid black; border-right: 2px solid black; padding: 2px; display: inline-block;"> 4P 2R </div>	4								
3	3				5		8		4	3	2				
4	4				4	7	3	2		1					
5	5					2	5	2	2	2	1				
6	6									1	1	0			
7	7											1	0	0	
8	8												1	0	0

Above line figures refer to P heads

Below line figures refer to R heads

system is narrow as is expected when $\omega_e \approx \omega_g$. The rapid change in intensity of the P and R branches in system $^31 \rightleftharpoons ^1\Sigma^+$ can be followed in this table. This change is associated with the change over of the direction of degradation of the heads and is another consequence of the nearness of the ω_e' and ω_g' values and of the B'' and B values. When B is of the same order as B''

CLASSIFICATION OF TIF BANDS

All the bands measured have been classified and are tabulated in the usual manner in Table V.

TABLE V—CLASSIFICATION OF TIF BANDS

<i>System</i> $^3I \rightleftharpoons ^1\Sigma^+$									
λ Å	I	ν cm. ⁻¹	O-C	ν' (ν'')	λ Å	I	ν cm. ⁻¹	O-C	ν' (ν'')
3105.1	0	32196		8, 13 R	2903.43		34432.0	0.1	4, 5
3090.2	0	32351		7, 12 R	2901.04	3	34460.2		4, 5 R
3065.22	0	32614.6		8, 12 R	2896.44		34515.0	-0.5	3, 4
3050.26	0	32774.5		7, 11 R	2891.54	4	34573.5		2, 3 P
3037.50	0	32912.2		6, 9 R	2890.64		34584.3	1.0	2, 3
3026.14	1	33035.8		5, 9 R	2886.03	6	34639.5		1, 2 P
3025.78	1	33039.8		8, 11 R	2885.57		34645.0	-1.0	1, 2
3022.67	1	33073.7		4, 8 P	2884.02	2	34663.9		6, 6 R
3015.23	2	33155.3		3, 7 P	2882.00	6	34687.9		0, 1 P
3010.02	1	33202.9		7, 10 R	2874.16		34782.6	1.1	5, 5
2998.25	1	33343.1		6, 8 R	2873.36	5	34792.3		5, 5 R
2987.33	2	33465.0		5, 8 R	2865.59		34886.6	-0.2	4, 4
2975.52	3	33597.8		3, 6 P	2863.82	7	34908.2		4, 4 R
2969.75	4	33663.0		2, 5 P	2858.53		34972.7	-2.0	3, 3
2964.01	4	33728.2		1, 4 P	2854.87	8	35017.7		3, 3 R
2959.68	1	33777.6		6, 8 R	2854.01		35028.1	1.0	2, 2
2950.24		33885.7	2.4	5, 7	2852.24	9	35049.9		2, 2 P
2948.79	2	33902.3		5, 7 R	2847.98	10	35111.0		1, 1 P
2942.31		33977.0	-1.3	4, 6	2847.27		35102.3	-1.0	1, 1
2938.39	2	34022.3		4, 6 R	2843.53	10	35157.3		0, 0 P
2936.19	4	34047.9		3, 5 P	2835.92	2	35239.4		5, 4 R
2935.23		34058.9	-0.7	3, 5	2828.36		35345.9	-0.8	4, 3
2929.90	4	34120.9		2, 4 P	2827.13	4	35361.2		4, 3 R
2929.39		34126.9	2.5	2, 4	2821.03		35437.7	-1.5	3, 2
2928.19	2	34140.9		2, 4 R	2818.44	5	35470.2		3, 2 R
2924.69	5	34181.7		1, 3 P	2814.80		35516.0	0.6	2, 1
2921.34	5	34220.8		0, 2 P	2810.63	7	35568.8		1, 0 P
2911.95		34331.2	-0.7	5, 6	2809.44		35583.8	-0.4	1, 0
2910.64	2	34346.6		5, 6 R					
<i>System</i> $^3O^+ \rightleftharpoons ^1\Sigma^+$									
2808.58	6	35594.7	-0.7	2, 4 Q	2737.55	9	36518.2	-0.7	2, 2 Q
2807.91	6	35603.2		2, 4 R	2737.10	9	36524.2		2, 2 R
2796.87	8	35743.7	0.1	1, 3 Q	2725.81	10	36675.5	0.8	1, 1 Q
2795.02	8	35767.4		1, 3 R	2724.86	10	36688.2		1, 1 R
2786.88	3	35871.8	0.8	0, 2 Q	2715.86	10	36809.8	0.1	0, 0 Q
2783.41	3	35916.6		0, 2 R	2713.66	10	36839.7		0, 0 R
2772.79	4	36054.2	-1.1	2, 3 Q	2702.82	2	36987.4	1.0	2, 1 Q
2772.20	4	36061.8		2, 3 R	2702.48	2	36992.1		2, 1 R
2761.09	7	36207.0	-0.2	1, 2 Q	2691.42	3	37144.1	-1.8	1, 0 Q
2759.75	7	36224.5		1, 2 R	2690.88	3	37151.6		1, 0 R
2751.00	6	36339.6	1.1	0, 1 Q	2681.89	0	37276.1		3, 1 R
2748.08	6	36378.3		0, 1 R	2668.79	1	37459.1		2, 0 R

TABLE V—(continued)

<i>System</i> $^1\Pi \leftarrow ^1\Sigma^+$							
λ Å	I	ν cm. ⁻¹	ν' (v")	λ Å	I	ν cm. ⁻¹	ν' (v")
2347.3	0	42588	1, 7	2244.31	5	44543.2	0, 2
2322.5	0	43044	1, 6	2227.00	8	44889.4	1, 2
2298.1	1	43501	1, 5	2221.00	9	45011.0	0, 1
2274.0	1	43961	1, 4	2203.94	8	45359.0	1, 1
2268.1	3	44078	0, 3	2198.02	10	45481.2	0, 0
2250.43	4	44422.2	1, 3				

The author wishes to express his thanks to Professor W. E. Curtis of Armstrong College, Newcastle, where this work was carried out, for his active interest in this problem and to the Department of Scientific and Industrial Research for a grant which enabled the work to be undertaken. He is also grateful to Professor A. C. Menzies of University College, Southampton, for the facilities placed at his disposal during the preparation of this paper.

SUMMARY

The spectrum of TlF has been investigated in absorption and also in emission by means of a high-frequency discharge. A doublet and a singlet system of bands have been found in absorption together with three continua all within the region 2000–3100 Å. The doublet system also occurs in emission. It is considered that the doublet system is due to $^3\text{O}^+ \rightleftharpoons ^1\Sigma^+$ and $^3\text{I} \rightleftharpoons ^1\Sigma^+$ transitions and the singlet system to a $^1\Pi \leftarrow ^1\Sigma^+$ transition. The following constants of these levels have been determined:

Level	ν_0	ω_0	$\tau_0 \omega_0$
$^1\Pi$	~ 45500	~ 360	—
$^3\text{O}^+$	36869.5	360.65	12.25
^3I	35180.7	439.87	8.60
$^1\Sigma^+$	0	475.00	1.89

The analysis of the absorption spectrum made by Boizova and Butkow is shown to be incomplete and incorrect. Predissociation occurs in all the band systems and is probably responsible for the band intensities being the same in both emission and absorption. A short discussion on the nature of the unstable levels causing predissociation is given. No sign of the Tl_{203} isotope has been detected.

REFERENCES

- Aston, F. W. 1932 *Proc. Roy. Soc. A*, **134**, 571.
 Boizova and Butkow 1936 *Phys. Z. Sowjet.* **5**, 765.
 Holst 1934 *Z. Phys.* **93**, 55.
 Howell, H. G. 1936 *Proc. Roy. Soc. A*, **155**, 141.
 Miescher and Wehrli 1934a *Helv. Phys. Acta*, **7**, 298.
 — 1934b *Helv. Phys. Acta*, **7**, 331.
 Rochester, G. D. 1936 *Proc. Roy. Soc. A*, **153**, 407.

Adsorption on Measured Surfaces of Vitreous Silica—II

BY W. G. PALMER, *St John's College, Cambridge*

(Communicated by Sir William Pope, F.R.S.—Received 7 January 1937)

In earlier experiments (Part I Palmer 1935) a method of estimating the adsorption area of powdered silica was developed, and absolute measurements were obtained of the adsorption at 25° of a variety of volatile substances, but the relative pressure of the vapour phase was confined to a range below 0.5. An account is now given of similar measurements in which the pressure was taken to saturation with the typical substances benzene, acetone and methyl alcohol.

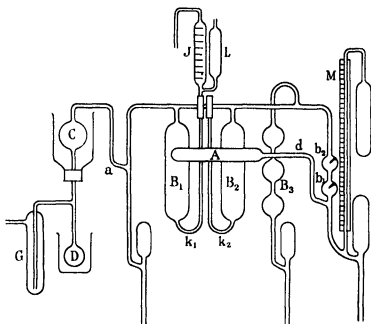


FIG. 1

Some modifications were made in the apparatus (shown in fig. 1 with lettering corresponding to Part I, fig. 1, p. 361) to cope with the larger adsorptions at higher pressures and the special difficulties in working with vapours near saturation. The superheating bulb *C* was immersed in water

kept at $70-80^\circ$ to ensure dryness in the vapour entering the storage bulbs B_1 and B_2 through the mercury trap a . The bulbs B_1 and B_2 were each of about 50 c.c. capacity and provided with capillary inlets k_1 and k_2 . The calibrated bulb B_3 (total volume 40 c.c.) and the upper manometer point b_2 were added so that, by allowing mercury to flow from some or all of B_3 , pressure in the storage system could be lowered and read with b_2 without affecting the pressure in the adsorption tube A . This was especially necessary in experiments with methyl alcohol whose vapour shows very marked imperfection near saturation. A large air thermostat kept the whole apparatus at $25 \pm 0.3^\circ$. As before, taps were completely excluded from the measuring parts of the apparatus.

DEVIATION OF THE VAPOURS FROM IDEAL BEHAVIOUR

Even if data were available for vapours at 25° it would still have been necessary to test the deviations in the actual apparatus to be used afterwards for adsorption. The connexion to the adsorption tube was sealed off at d , and after thorough exhaustion through G the inlet tubes k_1 and k_2 were sealed with a thread of mercury, the bulbs B_3 filled from below, and the mercury set to the point b_1 . Vapour was then taken from D through C and a until the relative pressure rose to about 0.5 when the cut-off a was closed. The pipette J was now charged *in vacuo* from L with freshly distilled mercury, which was brought to a temperature of 25° by an external heating wire. After atmospheric pressure had been restored in J a measured volume of warmed mercury was allowed to flow into B_1 and the new pressure read. The vapour was in this way compressed in easily controlled stages up to saturation, the bulb B_1 becoming nearly filled with mercury. The measurements were then repeated, the bulb B_2 alone being used. In all cases the two series of experiments led to identical correction values. Finally the capacity of the apparatus was calibrated with dry air by a similar method. If V_0 is the capacity and v the volume of mercury added to give a difference $p_1 - p_0$ in manometer readings, then for air

$$V_0 \cdot (p_1 - p_0)/v = p_1 + dp,$$

where dp is any systematic error in manometer readings.

A graph of $(p_1 - p_0)/v$ against p_1 gave an exact straight line passing through all the experimental points, and the slope of this at once gave V_0 . With benzene and acetone exact straight lines were also obtained to within 2 mm of saturation pressure, and the slope gave $V' = V_0 + a$. These vapours

therefore conformed accurately at all pressures to the well-known equation of state

$$V + a = \text{constant}/p$$

or

$$pV = p_i V(1 - a'p), \quad (1)$$

where p_i is the pressure of an ideal gas at volume V , and $a'p_0 = a/V_0$. Methyl alcohol vapour also obeyed this equation up to 115 mm., but between this and the saturation pressure at 124 mm. a marked divergence was observed.

TABLE I— $V_0 = 111.5$ C.C.

	a	a'
Acetone	2.4	0.000172
Benzene	4.7	0.000812
Methyl alcohol	6.6	0.000960 (to 115 mm.)

Corrected pressures were used in calculating adsorption. In the adsorption potential $\phi = -RT \ln f$, the relative fugacity was calculated for benzene and acetone from the relation (Lewis and Randall 1923, p. 198)

$$f = p^2/p_i,$$

but for methyl alcohol with its higher correction a third term in the expansion of f/p was taken into account (Lewis and Randall 1923). The assumption made in Part I that deviations from ideal behaviour could be neglected below relative pressure 0.5 is confirmed by the above observations except for alcohol. Fig. 2 (p. 261), which shows the results of recalculation from the original data for methyl alcohol, indicates that, as for other substances, an accurately linear relation exists between $\log \phi$ and the adsorption above low adsorptions, the abnormality of which was discussed.*

The fact that the vapours obey so closely the equation of state (1) is evidence that adsorption on the glass walls of the apparatus is negligible even at high pressures. To test this matter more directly vapour was expanded from the holders B_1 and B_2 (used in conjunction and also separately) into the empty adsorption tube attached at d as formerly. The expansion ratio by volume being exactly known, the final (corrected) pressure could be calculated and compared with that observed. Table II shows the results of a typical series of experiments with acetone.

Equally good agreement was found when only one storage bulb was used, so that no accidental compensation can have occurred.

* In the figures of Part I adsorptions per sq. m. were all printed without decimal points, and appear 10 times too large.

TABLE II—ACETONE

Saturation pressure 225 mm. Approximate apparent surfaces: storage bulbs 220 cm.²; adsorption tube 80 cm.²

p_1	Δp_1	Δp_2	Δp_2 calculated
19.6	19.6	15.3	15.3
43.1	27.8	21.7	21.7
76.6	39.6	31.1	31.0
112.1	44.0	34.4	34.4
143.6	41.1	32.2	32.1
183.5	48.8	38.1	38.15
Mercury at 25° admitted to B_1			
189.1	16.3	12.7	12.6
204.6	18.9	14.1	14.1
218.7	19.1	14.3	14.0
221.5	7.7	5.6	5.6

PARTICLE ANALYSIS OF THE SILICA POWDER

An analysis was undertaken by the usual method of sedimentation in water (Andreason 1929; Krauss 1927). The final fractions with r less than 9×10^{-5} cm. were separated by centrifuging at 60 G (see Table III).

TABLE III

Mean radius of fraction $\times 10^3$ cm.	Percentage	Contribution to specific surface cm. ²
355	43.86	(50)
300	18.08	84
200	16.14	109
126	9.11	97.5
82.3	4.85	80.0
58.5	3.26	75.5
41.2	2.49	82.5
28.9	1.54	71.0
17.5	1.48	115.0
9.1	0.56	84.0
6.1	0.106	24.0
4.9	0.049	13.6
3.1	0.021	9.0
		Total 895

In calculating column 3 of Table III the particles were assumed to be smooth spheres; the surface of the first fraction was estimated by microscopic examination. The actual specific surface was 3900 cm.² and the "roughness" factor is therefore 4.4. The mean radius r_0 of the equivalent uni-disperse powder is $3/(895 \times 2.2) = 1.5 \times 10^{-3}$ cm., and 78 % of the actual

powder consists of particles of radius greater than r_0 . It is evident from the above table that the surface of the powder is almost wholly made up of fairly uniform contributions from particles of radius between 3.0×10^{-3} and 1.0×10^{-4} cm., the contribution of particles of less size being negligible. Weight and bulk observations on the coarse unpowdered silica ($r_0 = 3.75 \times 10^{-3}$ cm.) and the finer material analysed showed that bulk/(vol. of silica) = 2.00 ± 0.05 in both cases, and hence the void space is equal to the nett volume of the material, and is therefore 3 times the void space for close-packed equal spheres. As in the latter assemblage the interstitial cross-section = $0.22 \times$ diameter of the spheres, we find for the mean interstitial cross-section in the powder

$$d_0 = 1.5 \times 0.44 \times \sqrt[3]{3} \times 10^{-3} = 0.95 \times 10^{-3} \text{ cm.}$$

ADSORPTION MEASUREMENTS

No changes were made in the method of preparing the silica, nor in the means of attachment to the apparatus.* About 10 g. of the powder were used, exposing approximately 5 sq. m. surface. The specimens of substances were from the stock formerly used. To estimate the surface by correlation with the earlier experiments adsorptions were first observed at three pressures within the range below 0.5. The close agreement (fig. 3) between the two series of observations made in entirely different apparatus is very satisfactory. Pressures up to about 0.7 were usually obtained by admitting fresh vapour from *D* (fig. 1) and higher pressures to saturation by compression with mercury as described above (p. 255). When, after the first series of experiments, the bulb B_1 was nearly filled with mercury, the silica was heated *in vacuo* and a second set of observations made in which only B_2 was used. Agreement within experimental error was found between the two series in all cases (fig. 3). With methyl alcohol vapour at pressures above 100 mm. the bulbs B_3 and the manometer point b_2 were brought into use, so that although adsorption pressures up to saturation were used, the adsorption was calculated from pressure measurements always below 100 mm. The isothermal of methyl alcohol falls so close to that of acetone that it could not be shown in fig. 3. At all pressures, even those nearest to saturation, adsorption equilibrium was reached at least as soon as the manometer reading could be taken, and several hours' interval, even with pressures within a few mm. of saturation, brought no further change.

* A detailed drawing of the mercury-sealed joint will be found in Part I, p. 361, fig. 1.

DISCUSSION

The following symbols will be used:

p_r = relative pressure

a = specific adsorption (micromol./sq. m.)

A = area per molecule (Ångström units)

ϕ = adsorption potential (cal./g. mol)

F = spreading or lateral force in the film (dynes/cm)

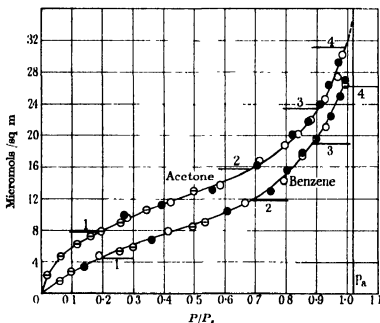


FIG. 3—○ experiments of Part I, ● experiments B_1 and B_2 , ○ experiments B_1 only.

1—The thickness of the adsorption film

Estimates of the apparent thickness of the film obviously depend on the area assumed to be occupied by 1 molecule of the adsorbate, which, owing to changes in orientation and packing, may not be constant at all stages in the film formation. Since, however, the areas found by the application of Langmuir's formula to experiments at low pressures (Part I, p. 381) corresponded in general with close-packing for small molecules (e.g. acetone, methyl acetate, acetonitrile), these areas were assumed to remain constant in calculating the thickness of acetone and methyl alcohol films in the present work. With benzene the low-pressure area was 36 Å^2 , implying an average tilt of the ring of 14° from the vertical; the close-packed area is

24 Å² (Adam 1930, p. 49; Mack 1925), and it seemed reasonable to assume that this area would be approached as the film thickened. As an approximation it was assumed that the outermost film would correspond with an area of 36 Å², and the inner layers with 24 Å². Table IV has thus been calculated, and indicates uniformity of behaviour in the three substances.

TABLE IV

	Relative pressure			
	1	2	3	4 layers
Methyl alcohol	—	0.61	0.80	1.0 (124 mm.)
Benzene	0.17	0.67	0.88	1.0 (94 mm.)
Acetone	0.19	0.68	0.90	1.0 (225 mm.)

Horizontal lines have been inserted in fig. 3 to mark apparent layers, but it must be emphasized that there is no indication in any part of the present work that layers are built up in regular succession, one being completed before the next is begun.

2—The form of the isothermals

The assumption that definite conclusions about the adsorption process and the complexity of the film can be drawn by considering the shape of the isothermal is falsified by much recent work. For example, de Boer (1931 *a, b*, 1932) for films of iodine on calcium and barium fluorides, and Frankenburger and Hodler (1932) for ammonia on tungsten, obtain S-shaped isothermals but prove the films to be not more than unimolecular throughout.

Schluter (1931) finds S-shaped curves for CS₂ on glass and silver surfaces, and curves wholly concave to the pressure axis for pentane on the same surfaces; the amounts adsorbed were approximately equivalent, and the films therefore of the same complexity in each case. Whipp (1933) obtains concave isothermals up to saturation pressure for iodine on potassium iodide, the films being 2 molecules thick at saturation. Cassel (1932) finds an S-shaped isothermal for CCl₄ on mercury, the film being 3.6 molecules thick at saturation. The following facts strongly support the assumption that purely adsorption forces are operative throughout the present experiments:

(a) The complete continuity of the linear relationship between $\log \phi$ and a from low pressures up to $p = 0.95$ (figs 4 and 5). Between this pressure and saturation the adsorption falls markedly *below* the expected values, and the isothermal intersects with the axis $p_r = 1.0$ at a finite value of da/dp (fig. 3). This cannot be due to the filling of interstices, as Coolidge suggested (1926), because these are so large (p. 258) that the corresponding adsorptions

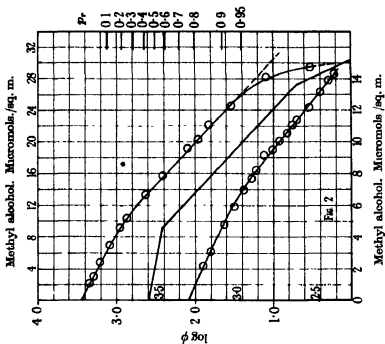


FIG. 5—(Fig. 2 inset)

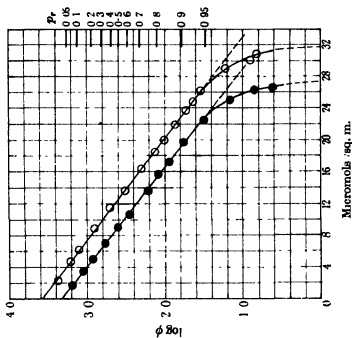


FIG. 4—○ acetone; ● benzene.

would be enormously greater than those observed. A more probable explanation is given later.

(b) Since the $\log \phi$ lines have a common slope, it is easily shown that the point of inflexion in the isothermals ($d^2p/da^2 = 0$) occurs for all three substances at the same relative pressure $= 1/e$. This fact, taken with the regularities in Table IV, is not compatible with any form of capillary condensation in view of the very different capillary properties of the substances (Floischer 1928)

(c) The adsorptions at saturation, and the amounts in excess of one close-packed layer (column 2) are nearly equimolecular (Table V):

TABLE V

	Micromols /sq. m.	
Benzene	28	21.3
Acetone	32	23.1
Methyl alcohol	30	22.0
Methyl acetate	32.5	25.5
Acetonitrile	29	22.9
Chloroform	33	—

By extrapolating
 $\log \phi$ lines from
 Part I

The differences from substance to substance are small enough to be reasonably attributed to different molecular area and possibly some difference in packing; they reveal no specific behaviour, but point rather to a process common to all the substances, viz. adsorption on the same surface.

The asymptotes p_a of the isothermals (fig. 3, acetone) are readily found by estimating the deficit $\Delta \phi$ in the potential near saturation.

TABLE VI

	$\Delta \phi_{(cal.)}$	p_a
Benzene	12	1.021
Acetone	10	1.016
Methyl alcohol	8	1.014

Thus it appears that liquid does not condense in bulk until the vapour phase is slightly supersaturated. The most probable reason for this lies in the high curvature of parts of the film surrounding sharp acute angles of the particles, which gives rise to a Kelvin effect and results in a slight thinning of the film at such points. For a surface tension of 30 units the rise of vapour pressure (from p_a above) would require a local curvature of radius $ca. 10^{-5}$ cm., which is not unreasonable on particles of the average diameter 1.5×10^{-3} cm. (p. 257).

3—Equations of state for the film

If the slope of the $\log \phi$ lines is s' , $-\log \phi/\phi_0 = s'a$ and $\phi = \phi_0 e^{-sa}$, where $s = 2.306 s'$, and $\log \phi_0$ is the intercept on the $\log \phi$ axis for $a = 0$. Also $A = 165/a$. We have also for a homogeneous surface

$$A dF = -kT d \ln f_r = d\phi_T.$$

On integrating for F we find

$$F = f\phi_0 e^{-sa}(a + 1/s) + I, \quad (2)$$

f being a numerical constant converting to appropriate units. In choosing the integration constant I we note that owing to the failure of the linear

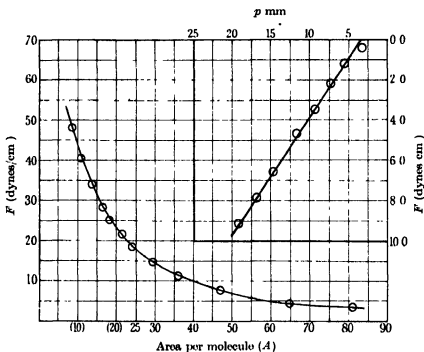


FIG. 6 Benzene

$\log \phi$ relation at very low adsorptions ϕ_0 does not correspond experimentally with $a = 0$ (Part I, fig. 10, p. 379). Inspection of the earlier data, especially for *c*-pentane, indicates that a linear relation is established when 9.0 % of a close-packed unimolecular layer has been adsorbed (a_1). It will be assumed provisionally that the bare surface remaining at this point is homogeneous, and in calculating F from (2), I is evaluated so that $F = 0$ when $a = 0.09a_1$ (cf. Langmuir 1932). A graph of $F-A$ for benzene is shown in fig. 6. It will be convenient to write values of A in brackets () when they

apply to films more than 1 molecule thick, and do not therefore represent the area per molecule in the direct sense.

a—Range of low adsorption ($A \leq 60$)—The rapid rise of the $FA-F$ curve for small values of F (fig. 7) and its marked convexity to the FA axis demand an equation of state of the form

$$(F - c/A^2)(A - A_0) = \text{constant} = i \cdot RT. \quad (3)$$

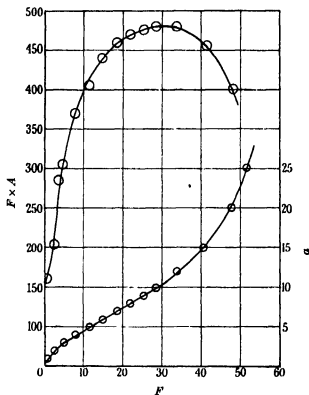


FIG. 7—Benzene.

Putting $i \cdot RT = 150$, $c = 10^4$, and $A_0 = 10$, we calculate F as follows:

TABLE VII—BENZENE

A	$c/A^2 = F'$	$i \cdot RT/(A - A_0)$	F_{calc}	$F_{\text{obs.}}$	μ'_D
200	0.25	0.80	1.05	1.00	1.68
100	1.00	1.60	2.6	2.60	1.41
80	1.60	2.10	3.7	3.70	1.35
70	2.05	2.4	4.45	4.40	1.31
60	2.80	3.0	5.8	5.60	1.26

F' is clearly a repulsive force, and $1/i$ measures the opposing attractive or cohesive force. If the force F' is attributed solely to dipoles induced in the adsorbate the corresponding dipole moment (μ' , column 6, Table VII) can be calculated from the equation given by Topping (1927). The permanent dipole moment in the surface (μ) requisite to induce this moment in benzene can be approximately found from the equation

$$\mu = \mu' \cdot c\alpha \cdot 3\epsilon r^3/(\epsilon + 2) = 3.0 D.$$

ϵ = dielectric constant of benzene = 2.3,

α = polarizability of benzene = 10^{-23} ,

r = distance between centres of dipoles = 3×10^{-8} ,

c = orientation factor, taken as 2.

Assuming further that the change in free-energy on adsorption for large values of A is due solely to dipole interaction between the surface and the adsorbate we find for the adsorption potential

$$\phi_0 = c^2 \alpha \mu^2 (\epsilon + 2)^2 / 9 \epsilon^2 r^6 = 1.92 \text{ ergs/mol} = 2750 \text{ cal./g. mol.}$$

Actually $\phi_0 = 2400 \text{ cal./g. mol.}$ In view of the approximate character of the calculation the agreement is no doubt fortuitous, but it serves to show that dipole interactions can account for the adsorption potential as well as the repulsive forces in the film. The dipole moments and cohesion constants in Table VIII are based on the assumption that a linear relation between $\log \phi$ and a is general, μ_0 is the dipole moment of the free molecule.

TABLE VIII

	μ	μ_0	i
Benzene	1.68	—	0.375
Chloroform	1.61	1.10	0.360
Acetone	2.10	2.80	0.600
Methyl acetate	2.35	1.75	0.725
Acetonitrile	2.20	3.20	0.640

The equation (3) reduces for large values of A to

$$F = \frac{\text{constant}}{A} (1 + bF). \quad (4)$$

A graph of p against F , calculated for benzene, assuming $I = 0$, shows a linear relation up to $F = 10$ dynes and $p = 20$ mm. (inset in fig. 6).

Putting $F = \text{constant}$ ($p - p_0$) = constant p' , we have from (4)

$$p'/(1 + bp') = \text{constant} \cdot A_1/A = \text{constant} \cdot \theta, \quad (5)$$

where A_1 is the area per molecule in a saturated unimolecular film. The equation (3) thus leads to Langmuir's formula, provided the pressure values

are adjusted to take account of the inhomogeneity of the surface. This was precisely the view taken empirically in Part I (p. 377 *et seq.*) where Langmuir's formula was shown to be applicable to low pressure portions of the isothermals. It is noteworthy that the connexion between (3) and (5) necessitates a positive sign in the bracket of (4), and therefore repulsive forces in the film.

b—Range of higher adsorption—Fig. 7 (lower graph) shows that for the range $a = 3$ to 12 and $A = 55$ to (14) the relation of F to a is linear, indicating an equation of state of the form

$$F = \text{constant} (1/A - 1/A_0), \quad (6)$$

or

$$F = \frac{\text{constant}}{A} (1 - b'A).$$

A_0 , the area at "zero compression", is *ca.* 100, which is in reasonable relation to the close-packed area of *ca.* 20. The range of validity of this equation is of interest because, as reference to fig. 3 shows, while the film is an expanded unimolecular one at the beginning of the range it is nearly 2 molecules thick at the end.

The general form of the FA - F curve (fig. 7) bears a resemblance to those for liquid expanded films on water as recorded by Adam (1926). Also Cary and Rideal (1925) found an equation of the form (6) fitted their data for the spreading of films from crystals on water. It seems probable that the formation of polymolecular films of small molecules on the solid corresponds with "condensation" of liquid expanded unimolecular films of long-chain molecules on water.

The assumption that 90 % of the silica surface can be regarded as homogeneous appears to be justified as it leads to a consistent scheme in analogy with films on water, and therefore the attempted explanation made in Part I, p. 383, of how Langmuir's formula might still apply to a supposedly heterogeneous surface is no longer necessary and is indeed at variance with the results recorded above. The inhomogeneity is probably not inherent in the surface, but arises from overlap of the surface force fields of contiguous particles in the powder.

SUMMARY

Earlier work on the adsorption of vapours on known surfaces of vitreous silica has been extended, and isothermals at 25° have been obtained from low pressures up to saturation for benzene, acetone, and methyl alcohol.

The S-shaped form of the isothermals is discussed, and reasons advanced for attributing the whole of the condensation to pure adsorption.

The films are 4 molecules thick at saturation, but the linear relation between the logarithm of the adsorption potential and the amount adsorbed holds strictly over almost the whole pressure range.

The spreading force is calculated from the experimental observations, and the values shown to agree with those calculable from certain equations of state for the films. Strong repulsive forces are accounted for quantitatively by dipole repulsion, and the observed maximum value of the adsorption potential can be successfully worked out on the basis of dipole interaction between the surface and the adsorbed molecules. The behaviour of the films of small molecules on the solid is compared with that of films of long-chain molecules on water.

REFERENCES

- Adam 1926 *Proc. Roy. Soc. A*, **112**, 377.
 — 1930 "Physics and Chemistry of Surfaces."
 Andreasson 1929 *Kolloidzechr* **48**, 175.
 Cary and Rideal 1925 *Proc. Roy. Soc. A*, **109**, 309.
 Cassel 1932 *Trans. Faraday Soc.* **28**, 178.
 Coolidge 1926 *J. Amer. Chem. Soc.* **48**, 1795.
 de Boer 1931a *Z. phys. Chem. B*, **14**, 149.
 — 1931b *Z. phys. Chem. B*, **15**, 281, 300.
 — 1932 *Z. phys. Chem. B*, **17**, 161.
 Fleischner 1928 *Amer. J. Sci.* **16**, 247.
 Frankonburger and Hodler 1932 *Trans. Faraday Soc.* **28**, 229.
 Krauss 1927 *Kolloidbeihfte*, **25**, 299.
 Langmuir 1932 *J. Amer. Chem. Soc.* **54**, 2798.
 Lewis and Randall 1923 "Thermodynamics."
 Mack 1925 *J. Amer. Chem. Soc.* **47**, 2468.
 Palmer 1935 *Proc. Roy. Soc. A*, **149**, 360.
 Schluter 1931 *Z. phys. Chem. A*, **153**, 68.
 Topping 1927 *Proc. Roy. Soc. A*, **114**, 68.
 Whipp 1933 *Proc. Roy. Soc. A*, **141**, 217.

Properties of Sufficiency and Statistical Tests

BY M. S. BARTLETT

*Statistician, Imperial Chemical Industries, Ltd., Jealott's Hill
Research Station, Bracknell, Berks*

(Communicated by R. H. Fowler, F.R.S.—Received 13 January 1937)

INTRODUCTION

1—In a previous paper*, dealing with the importance of properties of sufficiency in the statistical theory of small samples, attention was mainly confined to the theory of estimation. In the present paper the structure of small sample tests, whether these are related to problems of estimation and fiducial distributions, or are of the nature of tests of goodness of fit, is considered further.

The notation $a|b$ implies as before that the variate a is conditioned by† a given value of b . The fixed variate b may be denoted by $|b$, and analogously if b is clear from the context, $a|b$ may be written simply as $a|$. Corresponding to the idea of ancillary information introduced by Fisher for the case of a single unknown θ , where auxiliary statistics control the accuracy of our estimate, I have termed a conditional statistic of the form $T|$, quasi-sufficient, if its distribution satisfies the "sufficiency" property and contains all the information on θ . In the more general case of other unknowns, such a statistic may contain all the *available* information on θ .

SUFFICIENT STATISTICS AND FIDUCIAL DISTRIBUTIONS

2—It has been noted (Bartlett 1936a) that if our information on a population parameter θ can be confined to a single degree of freedom, a fiducial distribution for θ can be expected to follow, and possible sufficiency

* Bartlett (1936a): I have noticed an error on p. 128 of this paper which I will take this opportunity to correct. In the example the order of magnitude of the two observations was lost sight of. The information in one observation, *if it be recorded whether it is the greater or smaller*, is found to be 1.386, and is thus more than that in the mean.

† With this notation and phraseology, b is in general a known statistic. Inside a probability bracket, it may sometimes be necessary to stress that the distribution depends on an unknown parameter θ , and the usual notation is then adopted of writing $p(a)$ more fully as $p(a|\theta)$, and $p(a|b)$ as $p(a|b, \theta)$.

properties that would achieve this result have been enumerated. A corresponding classification of fiducial distributions is possible.

Since recently Fisher (1935) has put forward the idea of a simultaneous fiducial distribution, it is important to notice that the sufficient set of statistics \bar{x} and s^2 obtained from a sample drawn from a normal population (usual notation) do not at once determine fiducial distributions for the mean m and variance σ^2 . That for σ^2 follows at once from the relation

$$p(\bar{x}, s^2 | m, \sigma^2) = p(\bar{x} | m, \sigma^2) p(s^2 | \sigma^2), \quad (1)$$

but that for \bar{x} depends on the possibility of the alternative division

$$p\{\Sigma(x - m)^2 | \sigma^2\} p(t), \quad (2)$$

where t depends only on the unknown quantity m . No justification has yet been given that because the above relations are equivalent respectively to fiducial distributions denoted by $fp(m | \sigma^2)fp(\sigma^2)$ and $fp(\sigma^2 | m)fp(m)$, and hence symbolically to $fp(m, \sigma^2)$, that the idea of a simultaneous fiducial distribution, and hence by integration the fiducial distribution of either of the two parameters, is valid when both relations of form (1) and (2) do not exist (Bartlett 1936b). Moreover, even in the above example, the simultaneous distribution is only to be regarded as a symbolic one, for there is no reason to suppose that from it we may infer the fiducial distribution of, say, $m + \sigma$.

3—In certain cases where a fiducial distribution exists for a population parameter, it will similarly exist for the corresponding statistic in an unknown sample. If, for example, a sufficient statistic T_1 exists for θ in the known sample S_1 , we shall have a corresponding unknown statistic T_2 in an unknown sample S_2 , and an unknown statistic T for the joint sample S . If we write

$$p(T_1 | \theta) p(T_2 | \theta) = p(T | \theta) p(T_1, T_2 | T), \quad (3)$$

then $p(T_1, T_2 | T)$ depends only on T_1 and the unknown T_2 (for which T_1 may be regarded as a sufficient statistic), and will lead to a fiducial distribution for T_2 . Alternatively, if the unknown sample S_2 is merely the remainder of a "sample" from which, in order to infer its contents, a subsample S_1 has been drawn, we may obtain the fiducial distribution of T . If T_2 or T is an unbiased estimate of θ , we obtain the fiducial distribution of θ by letting the size of sample S_2 tend to infinity.

No corresponding fiducial distribution for T_2 (or T) exists if these statistics are only quasi-sufficient, since the configuration of the second sample will be unknown. T_2 has not then the same claim to summarize the contents of sample S_2 .

For similar inferences on both \bar{x} and s^2 (or \bar{x}_2 and s_2^2) in normal theory, the relevant probability distribution will be

$$p(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2 | \bar{x}, s^2), \quad (4)$$

which is necessarily independent of m and σ^2 . This distribution can be split up into two factors in three ways, corresponding to the association of s_1^2, s_2^2 or $(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2$ with the t -factor. We have

$$p\left(\frac{\bar{x}_1 - \bar{x}_2}{s_1}\right) p\left(\frac{s_2^2}{s^2}\right) \quad (5)$$

$$= p\left(\frac{\bar{x}_1 - \bar{x}_2}{s_2}\right) p\left(\frac{s_1^2}{s^2}\right) \quad (6)$$

$$= p\left(\frac{\bar{x}_1 - \bar{x}_2}{s}\right) p\left(\frac{s_1^2}{s_2^2}\right). \quad (7)$$

Since (5) is equivalent to $fp(\bar{x}_2)fp(s_2^2 | \bar{x}_2)$, and (7) to $fp(\bar{x}_2 | s_2^2)fp(s_2^2)$, it is consistent to speak of the simultaneous distribution $fp(\bar{x}_2, s_2^2)$. But while (5) is also equivalent to $fp(\bar{x})fp(s^2 | \bar{x})$, $fp(\bar{x} | s^2)$ is obtained from the first factor, and $fp(s^2)$ from the second factor, of (6), so that $fp(\bar{x}, s^2)$ also exists, but by virtue of a *different* factorization (cf. Fisher 1935).

For discontinuous variation, a relation (3) may similarly hold. While a fiducial distribution (in Fisher's sense) will no longer exist, the probability distribution $p(T_1, T_2 | T)$ will still be available for inferences on T_2 or T . Thus if S_1 contains r_1 members out of n_1 with an attribute A , etc., we obtain

$$\begin{aligned} p(r_1, r_2 | r) &= p(r_1)p(r_2)/p(r) \\ &= \frac{n_1! n_2! (n-r)! r!}{(n_1-r_1)! (n_2-r_2)! r_1! r_2!}, \end{aligned} \quad (8)$$

which will determine the chance, say, of obtaining as few as r_1 members in S_1 when S contains r such members, or S_2 at least r_2 such members.*

4—The equivalence of a sufficient statistic (or, when relevant, the fiducial distribution derived from it) to the original data implies that when it exists it should lead to the most efficient test. It does not follow that a uniformly most powerful test, as defined by Neyman and Pearson (1933), will necessarily exist, but if the probability (or fiducial probability) distribution is known, the consequences of any procedure based on it will also be known.

The converse principle, that the existence of a uniformly more powerful test must depend on the necessary sufficiency properties being present,

* For approximate methods of using equation (8), see Bartlett (1937).

and is, moreover, only possible for the testing of a single unknown parameter, has been denied by Neyman and Pearson (1936b); but while agreeing that the examples they give are formal exceptions, I think it is worth while examining their examples further, since they could reasonably be regarded as bearing out the principle to which formally they are anomalous. It seems to me more valuable to recognize the generality of the principle that a test of a single parameter should be most sensitive to variations in it than to reject the principle because of apparent exceptions.

In example I of their paper, the distribution

$$p(x) = \beta e^{-\beta(x-\gamma)}, \quad (x \geq \gamma), \quad (9)$$

is considered. It is required to test whether $\gamma \leq \gamma_0$ and/or $\beta \geq \beta_0$. Since if any observation occurs that is less than γ_0 no statistical test is necessary, we are effectively concerned only with samples for which all observations are greater than γ_0 . For such observations, the distribution law is

$$\begin{aligned} p(x) &= \beta e^{-\beta(x-\gamma)} / e^{-\beta(\gamma_0-\gamma)}, \quad (x \geq \gamma_0), \\ &= \beta e^{-\beta(x-\gamma_0)}, \end{aligned} \quad (10)$$

and is independent of γ . The sufficient statistic for β is \bar{x} , so that we are merely testing one independent parameter β for which a sufficient statistic \bar{x} exists.

Example II is merely a special case of this with $\beta = 1 + \gamma^2$, ($\gamma_0 = 0$), and again \bar{x} is the sufficient statistic for β and hence for γ , ($\gamma \leq 0$).

The above examples remind us, however, that a fiducial distribution is of more limited application than the sufficient statistic from which it is derived, and if restrictions are placed on the possible values of an unknown parameter, may become irrelevant. If the restriction is on an eliminated unknown, it might prove more profitable to use an inequality for a test of significance than an exact test. Thus if in normal theory it were known that $\sigma^2 \leq \sigma_0^2$, a test based on $p(\bar{x} | m, \sigma_0^2)$ might be more useful than one based on $p(t)$, though the possibility of using exact information on the range of other parameters in this way is in practice rare.

CONDITIONAL VARIATION AND EXACT TESTS OF SIGNIFICANCE

5—By exact tests will be meant tests depending on a known probability distribution, that is, independent of irrelevant unknown parameters. It is assumed that no certain information is available on the range of these extra parameters, so that their complete elimination from our distributions is desirable.

In order for the variation in the sample S to be independent of irrelevant unknowns ϕ , a sufficient set of statistics U must exist for ϕ . All exact tests of significance which are to be independent of ϕ must be based on the calculable *conditional variation* of the sample $S|U$. We may in fact state as a general principle that *all exact tests of composite hypotheses are equivalent to tests of simple hypotheses for conditional samples*. For this principle to be general, conditional variation is understood to include theoretical conditional variation; for we have seen that in certain cases allied to problems in estimation, the set U may be functions of a primary unknown θ .

A useful illustration of the principle is given by the known exact test (Fisher 1934, p. 99) for the 2×2 contingency table (observed frequencies $n_{11}, n_{12}, n_{21}, n_{22}$). The sufficient statistics U for the unknown probabilities of independent attributes A and B are $n_1 | n$ and $n_{.1} | n$, where $n_{.1} = n_{11} + n_{21}$, etc. Hence any exact test of independence *must* be based on the variation $S|U$, which has one degree of freedom, and a distribution

$$p(S|U) = p(S)/p(n_{.1} | n) p(n_1 | n) \\ = \frac{n_{.1}! n_{.2}! n_{.1}! n_{.2}!}{n_{11}! n_{12}! n_{21}! n_{22}! n!}. \quad (11)$$

6—It is of some importance to consider the relation of tests dependent on this principle of conditional variation with those obtained by the likelihood criterion introduced by Neyman and Pearson. Suppose there is only one degree of freedom after elimination of irrelevant unknowns, as in quasi-sufficient solutions of estimation problems, and suppose further the relation exists,

$$p(T_1 | \theta_1) p(T_2 | \theta_2) = p(T_1, T_2 | T) p(T | \theta), \quad (12)$$

when $\theta_1 = \theta_2 = \theta$. We have $p(T_1, T_2 | T) \equiv p(T_1 | T)$ and $T_1 | T$ is the statistic for testing discrepancies between θ_1 and θ_2 . By the likelihood criterion, however, the appropriate variate will be of the form

$$\lambda = \frac{p(T_1 | T) p(T | \hat{\theta})}{p(T_1 | \hat{\theta}_1) p(T_2 | \hat{\theta}_2)},$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ from the distribution of T , etc., whence

$$\lambda = \frac{f(T_1 | T) F(T)}{f_1(T_1) f_2(T_2)}, \quad (13)$$

say, which, for variation of S , will not necessarily be equivalent to $T_1 | T$, or independent of θ . The condition for λ to provide the same test as $T_1 | T$ appears to be that $T_1 | T$ should be equivalent to a function $\psi(T_1, T_2)$

independent of T , and that $F(T)/f_1(T_1)f_2(T_2)$ should be an appropriate function $\phi(\psi)$. This holds when λ provides the proper test in normal theory, but it clearly must fail when only quasi-sufficiency (not convertible into pure sufficiency) properties exist.

A modification of the criterion when statistics U exist is proposed here. For a comprehensive test of "goodness of fit" involving all the remaining degrees of freedom of the sample, the test may be based directly on the conditional likelihood of S |. For the joint testing of definite unknowns, the conditional likelihood of the relevant statistics T | would be considered. A criterion of this kind, if it differs from λ , is denoted subsequently by μ .

The mathematical definition of likelihood adopted is not separated from the more fundamental conception of the chance $p(S)$ of the observed data. For discontinuous data* the two are identical, so that the logarithm L is $\log p(S)$. For continuous variation it is sometimes convenient to drop the infinitesimal elements dx in $p(S)$, but some caution is necessary, this becoming more apparent when the likelihood of derived statistics is to be considered. Thus for s^2 (n degrees of freedom) in normal theory, $L(s^2)$ must, like $p(s^2)$, be invariant for a simultaneous change in scale in both s^2 and σ^2 , and is defined to be

$$C + n \log s - n \log \sigma - \frac{n}{2} \left(\frac{s^2}{\sigma^2} - 1 \right), \quad (14)$$

it being permissible to drop the term $d \log s^2$ (but not ds^2).

As an example we shall derive the μ appropriate for testing discrepancies among several variances. It is assumed that means and other regression parameters have already been eliminated, the statistic U , being the pooled residual variance s^2 , and the T the k individual variances s_r^2 . Then

$$L(T) = L(T | U) + L(U),$$

$$\text{or} \quad L(T | U) = C' + \sum n_r \log s_r - n \log s.$$

For convenience $L(T | U)$ is measured from its maximum value C' , so that

$$-2 \log \mu = n \log s^2 - \sum n_r \log s_r^2. \quad (15)$$

This criterion is not identical with that proposed by Neyman and Pearson, the weights being the number of degrees of freedom and not the number of observations. It is, however, put forward as an improvement, and a practicable method of using it derived below. With the original criterion λ it would be possible, if several regression parameters were eliminated from samples of unequal size, for fluctuations of a variance reduced to one or

* Variation in which only discrete values of the variate are possible is specified for convenience by the term "discontinuous variation".

two degrees of freedom to mask real discrepancies in more stable samples; this effect is corrected when the weight for such a variance is reduced correspondingly.

If any likelihood with f degrees of freedom tends to a limiting normal form $C \exp\{-\frac{1}{2}A(x, x)\}$, then $-2 \log \lambda$ will tend to be distributed as χ^2 with f degrees of freedom. This, apart from its practical importance, is a useful reminder of the goodness of fit, or test of homogeneity, character of such tests, and should warn us against pooling together components which there is reason to separate.

To obtain more precisely the value of the χ^2 approximation in the present problem, consider first μ from (14) (that is, $k=2$, $n_1=n$, $n_2=\infty$). From the known form (14) of the distribution of s^2 , we readily obtain by integration the characteristic function of $-2 \log \mu$ for this case; the expected value

$$M \equiv E(\mu)^{-2} = \frac{\Gamma\left(\frac{n}{2}(1-2t)\right) \left(\frac{n}{2}\right)^{nt} e^{-nt}}{(1-2t)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, \quad (16)$$

$$\begin{aligned} K \equiv \log M &= t \left(1 + \frac{1}{3n} + \dots\right) \\ &\quad + \frac{t^2}{2!} \left(2 + \frac{4}{3n} + \dots\right) + \frac{t^3}{3!} \left(8 + \frac{24}{3n} + \dots\right) + \dots \\ &= t \left(1 + \frac{1}{3n}\right) + 2 \frac{t^2}{2!} \left(1 + \frac{1}{3n}\right)^2 + 8 \frac{t^3}{3!} \left(1 + \frac{1}{3n}\right)^3 + \dots \end{aligned} \quad (17)$$

approximately. If we call the exact function $K(n)$, we have for the general equation (15), owing to the equivalence of the statistics $s_r^2 | s^2$ to angle variables independent of s^2 ,

$$K = \Sigma K(n_r) - K(n), \quad (18)$$

or neglecting the effect of terms of $O\left(\frac{1}{n_r^2}\right)$, as in (17), we write

$$-\frac{2 \log \mu}{C} = \chi^2 \quad (19)$$

with $k-1$ degrees of freedom, where

$$C = 1 + \frac{1}{3(k-1)} \left(\Sigma \frac{1}{n_r} - \frac{1}{n} \right). \quad (20)$$

The practical value of (19) was checked by means of the special case $k = 2$, representative values of (19) being given in Table I* for $P = 0.10$ and $P = 0.02$, corresponding to the correct values of χ^2 (one degree of freedom) of 2.706 and 5.412 respectively. The use of C tends to over-correct in very small samples the otherwise exaggerated significance levels, but it greatly increases the value of the approximation, and serves also as a gauge of its closeness.

TABLE I

n_1	$n_1 = n_2$					$n_2 = \infty$			
	1	2	3	6	12	2	4	9	∞
C	1.5	1.25	1.167	1.083	1.042	1.167	1.083	1.037	1
$P = 0.10$	2.48	2.66	2.69	2.70	2.70	2.68	2.70	2.70	2.706
$P = 0.02$	4.61	5.16	5.31	5.39	5.41	5.28	5.38	5.41	5.412

CONTINUOUS VARIATION—NORMAL THEORY

7—For continuous variation, such as in normal theory, exact tests of significance have often been obtained owing to the readiness of the sample to factorize into independent statistics. Thus, for all inferences on the normality of a sample the relevant distribution

$$p(S | \bar{x}, s^2)$$

is expressible as a product of t distributions, and the usual statistics g_1 and g_2 (or β_1 and β_2) for testing for non-normality are independent of \bar{x} and s^2 .

The usual χ^2 goodness of fit test is not, but since the expected frequencies corresponding to $p(S|)$ would in any case only be used in a "large sample" test, which is an approximation, the alternative use of the estimated normal distribution, $m = \bar{x}$, $\sigma^2 = s^2$, will be legitimate. We may appeal to Fisher's proof that the distribution of χ^2 when m and σ^2 are efficiently estimated follows the well-known form with $f-3$ degrees of freedom, where f is the number of cells.

But it is of theoretical interest to note that the *true* expected values for $S|$ could be found. Since the expected frequencies would then have three *fixed* linear conditions, for n , \bar{x} and s^2 , the number of degrees of freedom for χ^2 could, from this point of view, never have been questioned. $S|$ implies a sample point distributed over the $n-2$ -dimensional surface of a hyper-

* The values for $n_1 = n_2$ were obtained by means of Fisher's z table. When $n_2 = \infty$, the values for $n_1 = 2$ and 4 were obtained by an iterative method. It was subsequently noticed that the case $n_1 = 2$ could be checked from Table I of Neyman and Pearson's paper (1936 a), from which the values for $n_1 = 9$ were added.

sphere of radius $s\sqrt{(n-1)}$ in the "plane" $\Sigma x = n\bar{x}$, but the *expected* distribution of the n variates is more simply expressed by the distribution (n times that for any one variate)

$$E(S|) = \frac{n\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \left(1 - \frac{(x-\bar{x})^2}{(n-1)s^2}\right)^{\frac{1}{2}(n-2)} \frac{dx}{s\sqrt{(n-1)}}. \quad (21)$$

There is here a distinction between an exact test of goodness of fit for the normal law (which does not imply fitting a *normal* distribution at all), and the estimation of the normal law, which may be taken to be $m = \bar{x}$, $\sigma^2 = s^2$ (or $(n-1)s^2/n$).

Similarly for the exponential distribution with known origin 0 but unknown scale, the sufficient statistic for the scale is \bar{x} , the geometrical sample point is distributed at random over the "plane" $\Sigma x = n\bar{x}$, ($x > 0$), and the expected distribution is

$$E(S|) = \frac{n}{\bar{x}} \left(1 - \frac{x}{n\bar{x}}\right)^{n-2} dx. \quad (22)$$

For the distribution $p = dx/\pi\{1 + (x-m)^2\}$, for which

$$p(S|m) = p(\bar{x}|C, m)p(C) \quad (23)$$

(where C is the configuration), the goodness of fit will be based on C and the estimation problem has entirely disappeared.

8—For two correlated variates x_1 and x_2 , no function of the estimated correlation coefficients r and r' from two samples S and S' is a sufficient statistic for the true coefficient ρ . Hence no "best" test for a discrepancy in correlation coefficients is possible.

If, however, the degree of association between x_1 and x_2 were to be compared in the two samples on the basis of two populations of the same variability, an appropriate distribution is

$$p(v_{12}, v'_{12} | V_{11}, V_{12}, V_{22}), \quad (24)$$

where v_{12} is the sample covariance of S , V_{12} that of $S + S'$ (with elimination of both sample means), etc. This distribution, which is necessarily independent of σ_1^2 , σ_2^2 and ρ , is thus a valid test of the difference between two covariances, although owing to the conditional nature of the distribution, the test would be rather an impracticable one even if the mathematical form of the distribution were known.

DISCONTINUOUS VARIATION—POISSON AND BINOMIAL THEORY

9—For discontinuous variation, as for continuous variation which has been grouped, it is expedient for all but very small samples to be treated by approximate tests, but it is still important to consider the exact tests when they exist, not only for use with very small samples, but so that the basis of the approximate tests may be more clearly realized.

Consider first the Poisson distribution. For two samples with observed frequencies r_1 and r_2 , the distribution of $(r_1, r_2 | r)$, where $r = r_1 + r_2$, is simply a partition or configuration distribution giving the number of ways of dividing r between the two samples, and is

$$p(r_1, r_2 | r) = \frac{r!}{2^r r_1! r_2!}, \quad (25)$$

or the terms of the binomial distribution $(\frac{1}{2} + \frac{1}{2})^r$. This will be the distribution for testing a discrepancy between the observed values of two Poisson samples.

For several samples, we have similarly a distribution depending on the multinomial distribution

$$\left(\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}\right)^r. \quad (26)$$

The χ^2 test for dispersion is

$$\chi^2 = \sum (r_i - k_i)^2 / k_i = (n-1) k_2 / k_1 \quad (27)$$

(where the usual semi-invariant notation is used, so that $k_1 = r/n$). Before an exact test is possible, a suitable criterion must be adopted. χ^2 and the μ criterion will no longer be completely equivalent in small samples, but since for distributions near to the Poisson the ratio k_2/k_1 may be defined as an index of dispersion, it seems in practice convenient still to consider χ^2 , or equivalently (since k_1 is fixed and is the true expected value in (27)) the variance k_2 .

The moments of $k_2 | k_1$ are of some interest; they would be obtained from the identity

$$k_2 = \frac{1}{n-1} \{ \sum r_i^2 - n k_1^2 \}.$$

Thus, after some algebra with factorial moments, it is found that

$$\kappa_1(k_2 | k_1) = k_1, \quad (28)$$

$$\kappa_2(k_2 | k_1) = \frac{2k_1(nk_1 - 1)}{n(n-1)}. \quad (29)$$

These results may be compared with

$$\kappa_1(k_2) = m, \quad (30)$$

$$\kappa_2(k_2 - k_1) = \frac{2m^2}{(n-1)}, \quad (31)$$

and the approximate solution from (27),

$$\kappa_2(k_2 | k_1) \sim \frac{2k_1^2}{n-1}. \quad (32)$$

When a large number of samples are available, a Poisson distribution may be fitted. Analogously to the goodness of fit tests for the normal and other continuous distributions, the *true* expected frequencies for observed values 0, 1, 2, etc. could be found from the distribution of S , but as a good enough approximation we fit the estimated Poisson distribution, $m = k_1$. The true expected distribution for S is the binomial

$$n \left(\frac{n-1}{n} + \frac{1}{n} \right)^r. \quad (33)$$

10—Similar, though somewhat more complicated, properties hold for the binomial distribution. For two samples, the exact distribution for inferring the contents of a second sample was given by (8), and this distribution may similarly be used for comparing two known samples. The problem is a special case of a 2×2 contingency table where the marginal frequencies one way, corresponding to the sizes of the two samples, are absolutely fixed.

For l samples, we have similarly

$$p(S|) = \frac{(N-R)! R!}{N!} \prod_{i=1}^l \frac{n_i!}{(n_i - r_i)! r_i!} \quad (34)$$

(N and R referring to the total sample $S = \Sigma S_i$).

For l samples with more than a two-way classification of the contents, $R_1 \dots R_m$ being the total observed numbers in the m groups,

$$p(S|) = \frac{R_1! \dots R_m!}{N!} \prod_{i=1}^l \frac{n_i!}{r_{i1}! \dots r_{im}!}. \quad (35)$$

This corresponds also to an $l \times m$ contingency table.

For testing the dispersion among l binomial samples of equal size n , the usual test is

$$\chi^2 = \frac{N \Sigma (r_i - k_1)^2}{k_1(N - lk_1)} = \frac{N(l-1)k_2}{k_1(N - lk_1)}, \quad (36)$$

with $l-1$ degrees of freedom. The exact distribution of $k_2 | k_1$ could always be investigated, if necessary, from (34). The alternative use of the μ criterion is considered in section 12.

11—The moments of $k_2 | k_1$ could also be found*, using factorial moments of r_i . For example,

$$\kappa_1(k_2 | k_1) = \frac{nR(N-R)}{N(N-1)}, \quad (37)$$

where $k_1 = R/l$ (cf. equation (46), of which this is a special case).

It might be noticed that the factorial moment-generating function for (33) is either the coefficient of x^R in

$$\frac{(N-R)! R!}{N!} (1+x+xt)^{n_1} (1+x)^{n_2}, \quad (38)$$

or the coefficient of x^{n_1} in

$$\frac{n_1! n_2!}{N!} (1+x+xt)^R (1+x)^{N-R}. \quad (39)$$

The expression (39) is most readily generalized for classification of individuals into more than two groups, and becomes

$$\frac{n_1! n_2!}{N!} \prod_{j=1}^m (1+x+xt_j)^{R_j}, \quad (40)$$

while (38) is most easily generalized for the case of more than two samples, and becomes

$$\frac{(N-R)! R!}{N!} \prod_{i=1}^l (1+x+xt_i)^{n_i}. \quad (41)$$

The general case (35), corresponding to the $l \times m$ contingency table, is more complicated. Its generating function can be regarded as the coefficient of $x_1^{R_1} x_2^{R_2} \dots x_m^{R_m}$ in

$$\frac{R_1! R_2! \dots R_m!}{N!} \prod_{i=1}^l \left\{ \sum_{j=1}^m x_j (1+t_{ij}) \right\}^{n_i}. \quad (42)$$

For a large number of equal binomial samples, the expected distribution of $S |$ is the hypergeometric distribution

$$E(S |) = \frac{l \cdot n! (N-n)! R! (N-R)!}{(n-r)! r! (N-n-R+r)! (R-r)! N!}. \quad (43)$$

* Note added in proof, 23 March 1937. Compare Cochran's treatment (1936). The exact value for the variance κ_2 appears somewhat complicated, but for large l it becomes approximately $2n^2 R^2 (N-R)^2 / (N^4 (l-1) (n-1))$, which checks with Cochran's result.

12—The 2×2 contingency table has been already mentioned. The exact solution for testing interactions in a $2 \times 2 \times 2$ table is also known (Bartlett 1935), but the immediate derivation of the probability of any partition, complete with constant, is no longer possible, owing to the complication which lack of independence introduces. Thus the number of ways of filling up a 2×2 table is the coefficient of $x^{n_1} \cdot y^{n_2}$ in

$$(1 + x + y + xy)^n = (1 + x)^n (1 + y)^n$$

or
$$\frac{n_1! n_2! n_1! n_2!}{n! n!},$$

but the number of ways of filling up a $2 \times 2 \times 2$ table *when testing the interaction* is the coefficient of

$$x_{11}^{n_{11}} x_{12}^{n_{12}} x_{21}^{n_{21}} x_{22}^{n_{22}} y_{11}^{n_{11}} y_{12}^{n_{12}} y_{21}^{n_{21}} y_{22}^{n_{22}} z_{11}^{n_{11}} z_{12}^{n_{12}} z_{21}^{n_{21}} z_{22}^{n_{22}},$$

in $(x_{11}y_{11}z_{11} + x_{12}y_{12}z_{11} + x_{21}y_{11}z_{12} + x_{22}y_{12}z_{12} + x_{11}y_{21}z_{21} + x_{12}y_{22}z_{21} + x_{21}y_{21}z_{22} + x_{22}y_{22}z_{22})^n$, (44)

this last expression no longer factorizing.*

The expected value in the cell of a 2×2 table is the first factorial moment. For example,

$$E(n_{11}) = \frac{n_1 \cdot n_2}{n}. \quad (45)$$

While this result is evident, it should be noted that the expected values in other χ^2 problems have not remained unaltered when S is modified to $S|$, χ^2 for a contingency table appears to have a slight theoretical advantage here when approximations to small sample theory are being considered.

Since the expected value corresponding to (34) must also be expressible as a rational fraction, the solution in terms of a cubic equation (Bartlett 1935) is an approximation, valid for large sample theory.

For the $l \times m$ table the second factorial moment for any cell is

$$\frac{n_i \cdot (n_i - 1) n_{.j} (n_{.j} - 1)}{n(n-1)},$$

whence the expected value of χ^2 itself is readily shown to be

$$E(\chi^2) = \frac{n}{n-1} (l-1)(m-1), \quad (46)$$

* The symbols x_{11} , x_{12} , x_{21} and x_{22} represent four parallel edges of a cube, the y 's four other parallel edges, and the z 's the remaining four. Each observed frequency n_{ijk} , ($i, j, k = 1, 2$), corresponds to a corner of the cube, and hence to the three edges which intersect there. The sum of the frequencies at the end of every edge is fixed,

so that the bias of χ^2 is small and unimportant in comparison with the more general effects of discontinuity on its distribution (Yates 1934).

Since for an $l \times m$ contingency table to be tested for independence, $p(S|)$ is given by (35), the μ criterion is

$$\mu = \prod_{i=1}^l \prod_{j=1}^m \frac{n'_{ij}!}{n_{ij}!}, \quad (47)$$

where the n'_{ij} are the values of n_{ij} maximizing $p(S|)$. If this criterion were used, the values n'_{ij} must be found by inspection, though they will be near the expected values of n_{ij} . Equation (47) may be contrasted with the λ criterion given by Wilks (1935).

From (47) a small sample test is always possible. Thus for three binomial samples of 20 (equivalent to a 2×3 contingency table), with numbers

$$2 \quad 0 \quad 7$$

the exact significance level corresponding to μ (and also in this instance to χ^2) is found to be 0.007.

For large samples $-2 \log \mu$ will, like $-2 \log \lambda$, be distributed like χ^2 , the three tests becoming equivalent. For medium to small samples, for which an exact test is too laborious, it is doubtful whether the usual χ^2 test can be bettered, for μ is not easier to compute, and its approximate distribution is only known in so far as μ tends to the form $\exp(-\frac{1}{2}\chi^2)$.

13—If a test of significance is to be independent of the particular *kind of population* from which the sample values were obtained, the whole set S of sample values must be regarded as fixed. This might seem to imply that nothing is left to vary; but permutations of order are still possible. The relation of the sample values x with different groups or treatments, or with values of a second variate y , leads to tests for the significance of differences in means, or for the significance of apparent association. Thus in an experiment the assignment of treatments is at our choice; randomization ensures the validity of a test along these lines, this test tending to the usual test for a reasonable number of replications.

SUMMARY

Properties of sufficiency must necessarily be considered for all small sample tests of significance, whether these are related to problems of estimation and fiducial distributions, or are of the nature of tests of goodness of fit.

The idea of "conditional variation" is developed, and its bearing on common tests, depending either on continuous or discontinuous variation, is shown. In particular, the use of χ^2 and other likelihood criteria is re-examined; and a new application of χ^2 proposed for testing the homogeneity of a set of variances.

REFERENCES

- Bartlett 1935 *J.R. Statist. Soc. (Suppl.)*, 2, 248.
 — 1936a *Proc. Roy. Soc. A*, 154, 124.
 — 1936b *Proc. Camb. Phil. Soc.* 32, 560.
 — 1937 *J.R. Statist. Soc. (Suppl.)*, 4 (in the Press).
 Cochran 1936 *Ann. Eugen.* 7, 207.
 Fisher, R. A. 1934 "Statistical Methods for Research Workers," 5th ed.
 — 1935 *Ann. Eugen.* 6, 391.
 Neyman and Pearson 1933 *Philos. Trans. A*, 231, 289.
 — — 1936a *Statistical Research Memoirs*, 1, 1.
 — — 1936b *Statistical Research Memoirs*, 1, 113.
 Wilks 1935 *Ann. Math. Statist.* 6, 190.
 Yates 1934 *J.R. Statist. Soc. (Suppl.)*, 1, 217.

The Accurate Determination of the Freezing-points of Alloys, and a Study of Valency Effects in Certain Alloys of Silver

BY WILLIAM HUME-ROTHERY, *Warren Research Fellow of the Royal Society*
AND PETER WILLIAM REYNOLDS, *The Old Chemistry Department, Oxford*

(Communicated by Sir Harold Carpenter, F.R.S.—
Received 14 January 1937)

1—INTRODUCTION

In the alloys of silver with the four elements immediately following it in the Periodic Table, the silver rich alloys give rise to primary substitutional solid solutions for which the phase boundaries in the equilibrium diagram show well-defined valency effects. When the diagrams are plotted in atomic percentages, the liquidus and solidus curves fall more steeply with increasing valency of the solute, whilst the extent of the solid solution becomes less. It was shown (Hume-Rothery, Mabbott and Channel-Evans 1934) that, to a

first approximation, the liquidus curves followed a simple valency rule, such that alloys of identical equivalent composition* had identical freezing-points, and, hence, in dilute solutions where the liquidus curves were straight lines, the atomic depression of freezing-point was proportional to the valency of the solute. The experimental errors of 1 or 2° C. in the freezing-point data prevented an exact proof of a whole number law, and the present paper describes more accurate determinations of the freezing-points of these alloys. The first part of the experimental work was carried out by one of the authors (W. H.-R.) working alone, and the later work by the two authors together. It is convenient to refer to the two series of experiments as the early work and the later work respectively. In order to increase the accuracy of the liquidus determinations it has been necessary to examine many sources of error, and this work is described in § 2; § 3 contains the experimental results, which are discussed in § 4.

2—EXPERIMENTAL DETAILS

a—General Methods

In spite of its early date, the work of Heycock and Neville (1897) probably contains the most accurate determinations of the freezing-points of alloys previously made. These investigators used a resistance pyrometer, but in the present work it was desired to perfect a method suitable for use with small quantities of rare metals for which the heat capacity of a resistance thermometer is too great. The liquidus points have therefore been obtained by taking cooling curves in which the temperature was measured by a thermocouple contained in a very thin silica sheath. The general method was to melt the silver under a thick layer of powdered charcoal, and to determine its freezing-point in order to calibrate the thermocouple. The metal was then remelted, and the solute added with thorough stirring, after which the freezing-point of the alloy was determined. In the later work the alloy was stirred vigorously and continuously until the arrest point was well established, but in the early work the stirring had to be interrupted at each half minute while the temperature was measured and recorded.

In some cases successive additions of solute were made to the same melt ("method of successive additions"), but in others only one alloy was prepared from each portion of silver ("single alloy determinations"). In either case the thermocouple was very frequently standardized against the freezing-point of silver.

* By equivalent composition is meant the atomic percentage of solute multiplied by its valency.

b—Crucibles and Furnace Conditions

In the early work the alloys were melted under charcoal in cylindrical crucibles lined with fluorspar and alumina, the whole being contained in a larger crucible packed with charcoal. In the later work crucibles machined from a cylindrical graphite rod were used, and were much more satisfactory.

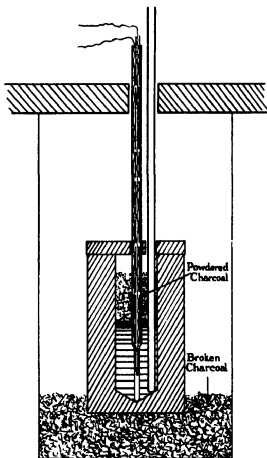


FIG. 1

The crucibles were provided with graphite lids through which holes were bored to admit a silica thermocouple sheath, a silica stirrer, and, where necessary, a silica tube by means of which a small portion of the melt could be removed for analysis. The general arrangement is shown in fig. 1. Provided the slabs at the top and bottom of the heating chamber were similar in a series of experiments, this arrangement gave the most reproducible

results, since it enabled the exact position of the crucible in the furnace, and particularly its depth, to be kept constant. With melts of 120–150 g., the effect of raising the thermocouple by 1–2 cm. altered the reading only by an amount of the order $0.1\text{--}0.2^\circ\text{C}$. The final arrangement was therefore to maintain the thermocouple at a constant depth, the end of the thermocouple being in a fixed position in the drawn-out tip of the sheath which was pushed to the bottom of the crucible and automatically centred by the conical depression (fig. 1). In the early work this precaution was not taken, but since the conditions were roughly constant in any one series, it is unlikely that any large error was introduced. For the highest accuracy it was essential to carry out all the cooling curves of a series, *including the silver-point calibrations*, under identical furnace conditions. Neglect of this precaution might easily introduce errors of the order $0.1\text{--}0.2^\circ\text{C}$., whilst errors of $0.5\text{--}1.0^\circ\text{C}$. occurred if the upper part of the crucible was excessively cooled owing to insufficient lagging at the top of the furnace.

c—Temperature Measurements

Platinum-platinum 13% rhodium thermocouples were used, since although their sensitivity is smaller than that of an alumel-chromel thermocouple, the reproducibility in the range $900\text{--}1000^\circ\text{C}$. is much greater. The cold junction was contained in a vacuum flask filled with a mixture of distilled water and pure ice prepared from distilled water, since slight errors may be introduced by the use of tap water and commercial ice. The general policy was to calibrate a thermocouple once against the freezing-points of silver (960.5°C .), the silver-copper eutectic (778.8°C .), and aluminium (659°C .), from which a deviation chart was constructed for use with the tables of Roeser and Wensel (1933). For the slight fluctuations of the standard points during the work, the deviation curves were so nearly parallel that no error greater than 0.05°C . was introduced by assuming them to be parallel, and in this way the temperature of each arrest point was deduced from the deviation graph and the silver point appropriate to the cooling curve concerned.

In the course of the work the final thermocouple was tested 38 times against the freezing-point of silver, and the extreme values of the electromotive force were 9.9681 and 9.9891 mV. The extreme difference of 0.021 mV is equivalent to 1.6°C ., and the cause of this variation was never satisfactorily established. It was shown conclusively that this source of error was not fortuitous, but that the observed electromotive force of the thermocouple underwent definite changes, and was usually consistent over short

b—Crucibles and Furnace Conditions

In the early work the alloys were melted under charcoal in cylindrical crucibles lined with fluorspar and alumina, the whole being contained in a larger crucible packed with charcoal. In the later work crucibles machined from a cylindrical graphite rod were used, and were much more satisfactory.

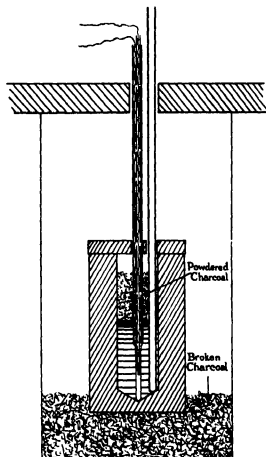


FIG. 1

The crucibles were provided with graphite lids through which holes were bored to admit a silica thermocouple sheath, a silica stirrer, and, where necessary, a silica tube by means of which a small portion of the melt could be removed for analysis. The general arrangement is shown in fig. 1. Provided the slabs at the top and bottom of the heating chamber were similar in a series of experiments, this arrangement gave the most reproducible

results, since it enabled the exact position of the crucible in the furnace, and particularly its depth, to be kept constant. With melts of 120–150 g., the effect of raising the thermocouple by 1–2 cm. altered the reading only by an amount of the order $0.1\text{--}0.2^\circ\text{C}$. The final arrangement was therefore to maintain the thermocouple at a constant depth, the end of the thermocouple being in a fixed position in the drawn-out tip of the sheath which was pushed to the bottom of the crucible and automatically centred by the conical depression (fig. 1). In the early work this precaution was not taken, but since the conditions were roughly constant in any one series, it is unlikely that any large error was introduced. For the highest accuracy it was essential to carry out all the cooling curves of a series, *including the silver-point calibrations*, under identical furnace conditions. Neglect of this precaution might easily introduce errors of the order $0.1\text{--}0.2^\circ\text{C}$., whilst errors of $0.5\text{--}1.0^\circ\text{C}$. occurred if the upper part of the crucible was excessively cooled owing to insufficient lagging at the top of the furnace.

c—Temperature Measurements

Platinum-platinum 13 % rhodium thermocouples were used, since although their sensitivity is smaller than that of an alumel-chromel thermocouple, the reproducibility in the range $900\text{--}1000^\circ\text{C}$. is much greater. The cold junction was contained in a vacuum flask filled with a mixture of distilled water and pure ice prepared from distilled water, since slight errors may be introduced by the use of tap water and commercial ice. The general policy was to calibrate a thermocouple once against the freezing-points of silver (960.5°C .), the silver-copper eutectic (778.8°C .), and aluminium (659°C .), from which a deviation chart was constructed for use with the tables of Roeser and Wensel (1933). For the slight fluctuations of the standard points during the work, the deviation curves were so nearly parallel that no error greater than 0.05°C . was introduced by assuming them to be parallel, and in this way the temperature of each arrest point was deduced from the deviation graph and the silver point appropriate to the cooling curve concerned.

In the course of the work the final thermocouple was tested 38 times against the freezing-point of silver, and the extreme values of the electromotive force were 9.9881 and 9.9891 mV. The extreme difference of 0.021 mV is equivalent to 1.6°C ., and the cause of this variation was never satisfactorily established. It was shown conclusively that this source of error was not fortuitous, but that the observed electromotive force of the thermocouple underwent definite changes, and was usually consistent over short

periods.* The most probable cause appeared to be accidental straining of the wires, which was gradually relieved on further heating. In the final part of the later work the greatest care was therefore taken to leave the thermocouple undisturbed between the silver-point calibrations and the cooling curves. The greatest change in the silver point at the beginning and end of a series was equivalent to 0.6°C ., and was usually very much less. Any slight change between the calibration points was assumed to have taken place in equal steps each time the thermocouple was heated. In the tables of results the estimated maximum error from this source is indicated,† but it must be emphasized that these are the maximum, and not the most probable errors. If, for example, the silver point changed by 0.3°C ., after heating three times, the temperatures were calculated on the assumption that the change took place in three steps of 0.1°C ., whilst the maximum estimated error in the tables is obtained by assuming that the whole change occurred suddenly so as to introduce the greatest possible error into the experiment concerned. The experiments suggest that the error from this source can be reduced to 0.1°C . when the thermocouple is checked against the silver point after every one or two curves, and results are automatically rejected when any marked change occurs. This frequent calibration is, however, essential, since calibrations at longer intervals may fail to reveal a change of half a degree which has persisted for a short time, and gradually died away. The above description refers to the relative accuracy of the results when compared with one another; the *absolute* accuracy depends on the tables of Roeser and Wensel, and any errors contained in these will be reproduced in the present work.

d—Potentiometer Adjustments

In the early work, an ordinary Carpenter-Stansfield potentiometer with steps of 0.5 mV was used, whilst in the later work the sensitivity was doubled by the addition of further resistance, and the coils corresponded with steps of 0.25 mV . By the kindness of the Aeronautical Research Committee, the instrument was calibrated at the National Physical Laboratory before and after the alteration; the errors removed by accurate calibration of the coils were of the order $0.1\text{--}0.3^{\circ}\text{C}$. With this instrument the accuracy is clearly greatest at temperatures for which the galvanometer deflexion is

* An alternative explanation was that the cadmium cell underwent variations owing to accidental overloading, but a detailed study did not support this view.

† The notation $\begin{pmatrix} +0.1 \\ -0.2 \end{pmatrix}$ means that a measured arrest point of 910.5°C ., for example, cannot, by virtue of changes in the thermocouple, represent a true liquidus point outside the limits 910.3 and 910.6°C .

small, since any slight inaccuracy in the sensitivity of the galvanometer produces a negligible effect near the zero point.* In the later work the compositions of the alloys were therefore chosen so that the arrest temperatures corresponded as nearly as possible with multiples of 0.25 mV, and errors from this source were reduced to less than 0.05° C. Further precautions were to check the zero point immediately after the arrest point, to test the sensitivity at each swing over from one coil to the next, and to secure an exact balance against the cadmium cell.†

e—Determination of the Composition of the Melt

In the alloys concerned, a change of 1 atomic % in composition alters the liquidus temperature by amounts varying from approximately 4° C. (cadmium) to 10° C. (antimony). A knowledge of the composition of the melt to an accuracy considerably greater than 0.1 % was therefore essential. Experiment showed that, with slowly cooled ingots, the segregation effects were often so great that the desired accuracy could only be obtained by dissolving the complete cooling curve ingot for analysis. Analysis of drillings from different parts of the ingot is not necessarily a satisfactory solution of this difficulty, particularly in ingots stirred during solidification, since the stirring may cause agglomeration of the first solid to separate. Apart from destruction of material, the complete dissolution of the cooling curve ingot leads to difficulties in the analytical technique, particularly if charcoal or flux is entrapped in the ingot. It is also inapplicable to alloys of volatile metals where changes in composition may occur on slow cooling of the ingot. The problem of determining the composition of the melt was examined in great detail, and the following conclusions have been reached.

(1) In alloys of common metals which are sufficiently‡ non-volatile, and can be melted without contamination, the most accurate results are obtained by taking the synthetic compositions of single alloy determinations.

(2) Where economy of material is important, the synthetic compositions may be used in conjunction with the method of successive additions, but

* The general arrangement with this kind of instrument is to adjust the sensitivity of the galvanometer so that a step of 0.25 mV corresponds with a simple number of scale divisions of the galvanometer. If on account of slight lack of adjustment the galvanometer reading is in error by 1.0 mm. at a deflexion of 10 cm., the error will be only 0.1 mm. at a deflexion of 1 cm. Errors due to the possible non-linearity of the scale at high angles of deflexion are also negligible near the zero point, so that no correction is necessary.

† If the resistance in the potentiometer circuit is adjusted only to the nearest 1 Ω , errors of 0.25° C. may be introduced.

‡ This point must be tested by confirming that no appreciable change in freezing-point occurs after relatively long heating of the alloy.

scrupulous care is necessary, since errors may be cumulative, and any slight loss by volatilization, oxidation, splashing, etc., affects all later compositions. If this method is used, the final ingot should therefore be weighed.

(3) In all other cases the best results are obtained by sucking up a small sample of the melt through a preheated refractory tube at a temperature slightly above the expected liquidus point. This method is simple and economical, and provided that the precautions described below are taken, it is satisfactory for liquidus determinations where an accuracy of ± 1 or $\pm 2^\circ \text{C}$. is sufficient. For more accurate work unexpected difficulties arise with some alloys, and these have been fully examined. For this purpose samples of our silver alloys were removed by gentle suction through a preheated silica tube of $\frac{1}{4}$ in. internal diameter. The preheating of the tube is essential, since if a cold tube is plunged into a molten alloy, partial solidification may occur, and the differences in composition of the solid and liquid phases may produce quite misleading results. The suction tube must not, however, be kept in the alloy for a long time before the sample is removed, since the portion of the alloy within the suction tube is then isolated from the main bulk of the alloy, and if the composition is gradually changing owing to volatilization, the conditions for the two parts are not the same.

With silver-indium alloys the non-volatility of the indium enabled the sampling method to be tested against the synthetic compositions of single alloy determinations. With alloys containing 6.00, 9.19 and 11.75 % of indium, the indium contents determined by analysis of samples differed from the synthetic compositions by -0.01 , $+0.02$ and $+0.01$ % respectively. The sampling method was here entirely satisfactory. On the other hand, with silver-tin alloys, inconsistent results were obtained, and in nearly all cases the tin contents of the extracted samples were greater than those expected from the synthetic compositions by amounts of the order 0.05–0.15 %. The high tin figures were not due to analytical errors, since the sums of the percentages of silver and tin determined by analysis lay between the limits 99.96 and 100.006, and the method used was one in which the silver was completely removed by precipitation before the tin was determined. The whole problem has been examined in detail, and the most probable explanation appears to be that as the suction tube is raised, solidification of the top of the sample begins, and produces segregation effects. Consequently if, as normally happens, a small drop of liquid or semi-liquid alloy falls back into the crucible when the bottom of the suction tube leaves the surface of the main alloy, the composition of the sample may no longer be exactly that of the melt. This explanation was supported by the fact that whilst in some cases the composition of the extracted sample was uniform

over its whole length, in other cases differences of the order 0.2–0.3 % were found between the compositions of different parts. It was not found possible to prevent this effect by constricting the bottom of the suction tube. These errors are clearly diminished by removing as large a sample as is possible with a thin suction tube, after which the whole sample must be used for analysis. The errors from this source may be expected to increase with increasing distance between the solidus and liquidus curves, but this is not the only factor concerned, since although the same effect was found for silver-antimony alloys, it was much less pronounced than with the silver-tin alloys. It is perhaps significant that the latter were the only alloys for which we found longitudinal segregation in the rods used for our previous work on lattice spacings of silver alloys.

f—Errors due to Supercooling

The well-known tendency of silver alloys to show supercooling effects made it necessary to examine the errors involved, and for this purpose use was made of silver-tin and silver-indium alloys which could be remelted without change in composition. Supercooling could not be prevented by seeding the melt, or by merely reducing the rate of cooling. The best results were obtained by using a special method of stirring, together with a very slow rate of cooling. The alloy was stirred by moving the stirring rod rapidly up and down through a distance of about 1 cm. At intervals of every 15 sec., the rod was lifted up once about 3 cm., and the rapid stirring then continued. By repeating this process at exactly regular intervals, the rate of cooling, as determined by readings every 30 sec., was quite uniform, but the slight additional chilling produced by raising the stirrer through the higher distance was sufficient to start the crystallization. Fig. 2, curve *A*, shows a cooling curve obtained in this way with an alloy containing 8.5 atomic % of tin. It must be emphasized that this is the correct form of cooling curve for an alloy which freezes over a range of temperature. In the absence of supercooling, crystallization should set in when the liquidus point is reached, and the rate of cooling should diminish, but there should be no sudden shower of crystals, or horizontal arrest point. Fig. 2, curve *B*, shows a cooling curve of the same alloy taken, without stirring, at a rate of 1.1° C./min. In this the alloy has supercooled 1.5° C., below the true liquidus point, but there is no appreciable recalescence,* presumably because the freezing began at

* For convenience the term recalescence is used to denote the actual rise of temperature observed after freezing begins, and supercooling to denote the extent to which the temperature falls below the true liquidus point before solidification begins.

the outside of the ingot so that, *in the absence of stirring*, the heat liberated was conducted away from the thermocouple.* Fig. 2, curve C, shows the cooling curve for the same alloy which was allowed to cool, without stirring, well below the liquidus temperature, when vigorous stirring was begun. Here the alloy has supercooled 1.0°C . below the true arrest point, but owing to the thorough mixing produced by the stirring there is now 0.45°C . of recalescence, and the maximum on the curve is only 0.5°C . below the true arrest point. The errors from this source clearly increase with (a) the rate of cooling, (b) the steepness with which the liquidus curve falls, and (c) the

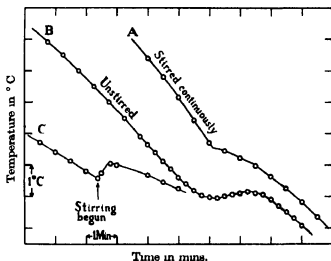


FIG. 2

distance between the solidus and liquidus curves, since the greater this distance the smaller the amount of solid which crystallizes out for a given fall in temperature, and hence the smaller the amount of latent heat available to raise the temperature. Table I shows results for an alloy containing 2.67 atomic % of indium. In this case it was never possible to obtain an

* It must be emphasized that the fact that there is no observed recalescence, or horizontal portion of the cooling curve, does not necessarily denote a satisfactory arrest point. If, for example, the top of the furnace is insufficiently lagged, solidification will begin at the unduly cooled surface of the melt, and any supercooling of this portion may not be detected by the thermocouple below. Under these circumstances, the apparent arrest point may be above the true liquidus point, since the first observed retardation of cooling is due to the lagging effect of the solidifying portion, and does not correspond with the true liquidus point of the melt in contact with the thermocouple. Thorough stirring and uniform cooling are therefore essential for accurate results, and temperature gradients in the furnace should be reduced as much as possible.

arrest completely free from recalescence, but by reducing the rate of cooling to 0.5° C./min. the recalescence was reduced to 0.1° C. In this case the maximum on the curve was at 944.1° C., and was only 0.3° C. higher than the value obtained with a recalescence of 0.7° C. at a rate of cooling three times as great. The thermocouple was in a very steady state during this work, and the thermocouple errors should not exceed $\pm 0.05^\circ \text{C}$.

TABLE I—SILVER-INDIUM ALLOY CONTAINING 2.67 ATOMIC % INDIUM

Rate of cooling °C./min.	Recalescence °C.	Maximum temperature on cooling curve after the beginning of the arrest °C.
1.5	0.7	943.8
0.9	0.45	943.9
1.0	0.35	944.0
0.5	0.1	944.1

In the silver alloys dealt with in the present paper the distance between the liquidus and solidus curves, and the steepness of the liquidus curve increase with the valency of the solute, so that errors due to supercooling become more serious as the valency of the solute increases. Fortunately, with one exception, the critical points for the silver-antimony alloys are from curves which show little or no recalescence. The data given above suggest that with thoroughly stirred alloys under our experimental conditions, cooling curves showing one- or two-tenths of a degree of recalescence will only introduce errors of one- or two-tenths of a degree for silver-rich alloys, increasing to not more than 0.5° C. for the most concentrated alloys concerned in this investigation. It must be emphasized that this applies only to experiments with a very thin thermocouple sheath in a well stirred alloy. With a thick thermocouple sheath the problem is much more complicated owing to the heat capacity of the sheath, and the lag between the temperature of the thermocouple and the alloy.

g—Errors due to Volatilization of one Constituent

For the silver-cadmium alloys a correction had to be made for the change in composition of the alloy which occurred between the extraction of the sample for analysis, and the arrest point. For this purpose four silver-cadmium alloys were prepared with compositions such that their liquidus temperatures corresponded as nearly as possible with multiples of 0.25 mV of the thermocouple (see p. 287). For each alloy a cooling curve was taken until the arrest point was established and the time was noted. The tempera-

ture was then raised immediately and the alloy maintained slightly above the arrest point for a period of from $\frac{1}{2}$ to $1\frac{1}{2}$ hr. (according to the composition), and was stirred for the whole of this period in the same manner as during the taking of a cooling curve. A second cooling curve was then taken, and the time of the arrest point again noted. In this way the rise of liquidus temperature produced by the volatilization of cadmium during a known period under the particular experimental conditions was obtained. It was assumed that the effect of volatilization was proportional to the time, and in this way the graph shown in fig. 3 was constructed to show the rise in

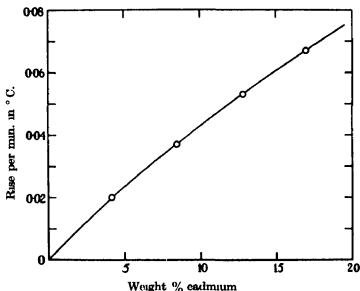


FIG. 3—Showing, as a function of the composition of the alloy, the rise per minute in liquidus temperature produced in silver-cadmium alloys heated just above the freezing-point with continuous stirring.

liquidus temperature per minute of heating slightly above the freezing-point, as a function of the composition of the alloy. From this curve the appropriate corrections were made to each observed liquidus point, according to the time which elapsed between the withdrawal of the sample and the arrest point.

3—EXPERIMENTAL RESULTS

The silver used throughout the present work was assay silver grain of at least 99.99 % purity obtained from Messrs Johnson, Matthey and Co., Ltd.

a—Silver-Cadmium

The cadmium used in the early work was specially pure metal kindly presented by the Metropolitan Vickers Electrical Co., Ltd., whilst for the later work use was made of spectroscopically pure cadmium kindly presented by the New Jersey Zinc Corporation. The results obtained are given in Table II, which includes the observed arrest points, the times elapsing between the withdrawal of the samples for analysis and the arrest points, and the temperatures corrected to allow for the volatilization of cadmium. In the course of the early work four points were obtained in the range

TABLE II*—LIQUIDUS POINTS FOR SILVER-CADMIUM ALLOYS

No.	Wt. % Cd	Atomic % Cd	Arrest points in °C.		Thermo- couple error (max.) °C.	Reca- lescence in °C.	Rate of cooling °C./min.	Interval between sampling and arrest point in min.
			Observed	Corrected for volati- lization				
1	3.61	3.47	945.9	945.8	(+0.05) (-0.15)	0.0	0.7	6
2	4.23	4.07	944.3	944.2	±0.0	0.0	1.0	4
3	4.24	4.08	943.8	943.7	(+0.15) (-0.1)	0.4	1.6	6
4	8.46	8.15	926.0	925.9	(+0.1) (-0.2)	0.2	1.5	3
5	9.51	9.16	922.3	921.5	(+0.15) (-0.05)	0.15	0.6	19
6	12.53	12.08	909.6	909.5	±0.0	0.1	0.9	2½
7	13.21	12.75	906.5	906.3	±0.0	0.0	1.4	4
8	13.30	12.83	905.8	905.1	±0.0	0.0	0.3	13
9	14.73	14.22	900.7	900.1	±0.0	0.0	0.8	11
10	16.76	16.20	890.2	890.1	(+0.15) (-0.2)	0.0	1.0	1

* The analyses were made by Messrs Johnson, Matthey and Co., Ltd.

920–950° C. When compared with the final results, one of these points is too high by 1.4° C., but the remaining points agree with the final curve within the limits $\pm 0.5^\circ$ C. and ± 0.05 atomic % on the composition axis.† These points therefore confirm the final results, but we have omitted them from the tables in order that all the results may be discussed to the same degree of accuracy. The present results agree with those in the collected tables of Hume-Rothery, Mabbott, and Channel-Evans (1934) to within

† In this work the time elapsing between the withdrawal of the sample and the arrest point was not noted, but any correction for this error would make the agreement even more exact.

1 or 2° C., but the experimental points in the earlier work were much less regular.

b—Silver-Indium

The indium used was obtained from the Indium Corporation of America, and was refined by Messrs Johnson, Matthey and Co., Ltd.; after the refining process it contained no impurities which could be detected by chemical means. The results are given in Table III in which the compositions are the

TABLE III*—LIQUIDUS POINTS FOR SILVER-INDIUM ALLOYS

No.	Wt. % In	Atomic % In	Arrest point °C.	Thermo-couple error (max.) °C.	Recalescence °C.	Rate of cooling °C./min.
1a	2.839 (S)	2.673	943.9	(+0.0) (-0.1)	0.45	0.9
1b	2.839 (S)	2.673	943.8	± 0.05	0.7	1.5
1c	2.839 (S)	2.673	944.0	± 0.05	0.35	1.0
1d	2.839 (S)	2.673	944.1	(+0.1) (-0.0)	0.1	0.5
2	2.840 (S)	2.674	943.8	(+0.1) (-0.05)	0.25	1.4
3	5.996 (S)	5.656	924.6	± 0.1	0.05	0.2
4	9.187 (S)	8.684	904.2	± 0.0	0.15	0.5
5	11.749 (S)	11.123	886.8	± 0.2	0.0	0.4

* (S) against the composition indicates that this was determined by synthesis.

synthetic compositions of single alloy determinations, and, as previously stated, the compositions of alloys Nos. 3, 4 and 5 as determined by analysis of samples extracted from the melts agreed with the synthetic values to within 0.02 %. Nos. 1a, 1b, 1c, and 1d are the same alloy remelted four times, but alloy No. 2 of nearly the same composition is an entirely different melt. The results for alloys Nos. 3, 4, and 5 were confirmed by an earlier series of almost the same compositions, but at this stage of the work the methods of sampling and analysis had not been perfected†, and as the experimental error was much greater, the results are not included. The liquidus points given by Hume-Rothery, Mabbott, and Channel-Evans (1934) for alloys containing 3.0, 8.0, and 10.0 atomic % of indium are in good agreement with those of the present work in spite of the fact that melts of only 10 g. were used, whilst their point for an alloy containing 4.95 atomic % of indium, which showed slight recalescence, is

† The synthetic composition could not be determined here because the method of successive additions was used and some of the alloy was split.

from 3 to 4° C. too low. The results of Weibke and Eggers (1935), who used melts of 20-25 g., are not in such good agreement, since the points for their third and fourth alloys containing 5.0 and 8.2 atomic % indium, are from 7 to 8° C. lower than the present values.

c—Silver-Tin

"Chempur" tin was used for this work, and the results are given in Table IV, the compositions being the synthetic values for single alloy determinations. The results were confirmed by further series in the early and the later work in which the compositions were taken as the synthetic values in the method of successive additions. These values are not included in the table because the error was greater owing to the fact that samples

TABLE IV—LIQUIDUS POINTS FOR SILVER-TIN ALLOYS

No.	Wt. % Sn	Atomic % Sn	Arrest point °C.	Thermo-couple error (max.) °C.	Recalescence °C.	Rate of cooling °C./min
1	2.470 (S)	2.250	943.8	± 0.0	0.0	2.0
2	2.474 (S)	2.254	943.9	± 0.2	0.1	0.5
3	2.476 (S)	2.255	943.6	± 0.0	0.4	0.8
4a	5.010 (S)	4.574	925.0	(+ 0.1) (- 0.2)	0.3	0.9
4b	5.010 (S)	4.574	924.9	(+ 0.2) (- 0.1)	0.05	0.3
5a	5.080 (S)	4.638	924.2	± 0.0	0.55	1.0
5b	5.080 (S)	4.638	924.2	± 0.0	0.4	0.5
5c	5.080 (S)	4.638	924.3	± 0.0	0.0	0.5
6a	7.135 (S)	6.527	907.3	(+ 0.1) (- 0.2)	0.0	0.6
6b	7.135 (S)	6.527	907.2	(+ 0.2) (- 0.1)	0.05	0.6
7a	9.230 (S)	8.460	888.4	± 0.3	0.0	0.6
7b	9.230 (S)	8.460	888.8	± 0.15	0.2	0.6
7c	9.230 (S)	8.460	888.4	± 0.05	0.1	0.5

were extracted for analysis, and the final weight was not checked. As previously explained, attempts to determine the composition from the analysis of extracted samples led to inconsistent results, but all the results based on synthetic compositions are in good agreement. Accurate determinations for this system were made by Heycock and Neville (1897) whose silver-tin series was headed by a value of 959.2° C. for the freezing-point of silver. If it is assumed that this was due to a constant error in the pyrometer used, the figures of Heycock and Neville would have to be raised by 1.3° C. in order to compare with the present results, and in this case their

values would be about 1° C. higher than those of the present work. This discrepancy would be accounted for if Heycock and Neville's low value for the freezing-point of silver were in this case† due partly to slight oxidation.

d—Silver-Antimony

The antimony used was kindly presented by The Cookson Lead and Antimony Co., Ltd., and was of 99.917 % purity, the chief impurities being iron 0.041 %, lead 0.025 %, and copper 0.012 %. The results obtained are shown in Table V, the compositions by synthesis of single alloy determinations being marked (*S*) in order to distinguish them from those obtained by the analysis of extracted samples. The liquidus points based on synthetic

TABLE V—LIQUIDUS POINTS FOR SILVER-ANTIMONY ALLOYS

No.	Wt. % Sb	Atomic % Sb	Arrest point °C.	Thermo-couple error (max.) °C.	Re-calescence °C.	Rate of cooling °C./min.
1	1.15	1.02	951.4	± 0*	0.0	1.7
2	2.159 (<i>S</i>)	1.918	942.7	± 0.0	0.0	0.3
3	2.28	2.03	942.3	± 0*	0.0	1.2
4	2.81	2.50	937.1	± 0*	0.15	2.0
5	3.35	2.98	932.5	± 0*	0.2	3.0
6	3.94	3.51	927.6	± 0*	0.0	2.0
7 _a	4.206 (<i>S</i>)	3.744	924.0	± 0.0	0.35	0.5
7 _b	4.206 (<i>S</i>)	3.744	924.1	± 0.0	0.45	0.5
8	5.06	4.51	916.4	± 0*	0.0	1.8
9	6.108 (<i>S</i>)	5.449	905.1	± 0.0	0.0	0.7
10	7.404 (<i>S</i>)	6.616	889.8	± 0.2	0.0	0.8

* In these cases the thermocouple was standardized at less frequent intervals, so that although no change in the silver point was detected, the errors may be greater than those stated.

compositions are very slightly below those based on analysis, and as this difference is in the opposite direction to that which would be expected if volatilization occurred, it appears that the sampling method is again slightly affected by segregation errors,‡ although the difference is only of the order 0.05–0.10 %. The points based on synthetic composition are to be preferred for critical discussion. The liquidus points of Heycock and Neville (1897) for this system are headed by a value of 958.8° C. for the freezing-point of

† Heycock and Neville's values for the freezing-point of silver vary from 958.8 to 962.3° C., but as pointed out previously (Hume-Rothery, Mabbott, and Channel-Evans 1934) a study of the curves suggests that most of the difference was constant in any one series of alloys, and not due to oxidation of the silver.

‡ Both silver and antimony were determined by one of us (Reynolds and Hume-Rothery 1936), and the analysis appeared accurate to within ± 0.03 to ± 0.05 %.

silver, and if they are raised by 1.7°C . they are about 0.5°C . above those of the present work.

4—DISCUSSION

The following points must be remembered in comparing the results for the different systems. (1) At a given temperature the tendency to obtain low results owing to undetected supercooling increases with the valency of the solute, since this results in a steeper fall in the liquidus curve, and a greater distance between the solidus and liquidus. (2) From the evidence obtained in the silver-tin and silver-antimony series (see p. 288) any errors due to difference between the composition of the extracted sample and that of the melt will probably tend to make the silver-cadmium points too high since the sampling method had to be used for this series.

The most reliable experimental results are shown in fig. 4, in which the liquidus points are plotted against the equivalent compositions of the alloys.* For the silver-antimony alloys only the results based on synthetic compositions are included, since the sampling method is undoubtedly slightly less accurate. In some cases where duplicate experiments gave results too close together to be distinguished on this scale, only one point is included. It must be emphasized that although this figure only shows 23 experimental points, these are confirmed by a large number of other results for which the accuracy is only slightly less, and which are omitted in order to avoid discussing individual points to different degrees of accuracy.

From fig. 4 it will be seen that the points for the silver-cadmium and silver-indium alloys lie on one curve for at least the first 40°C . depression of freezing-point, whilst the points for the silver-tin and silver-antimony alloys lie on a second curve for the whole range $960.5\text{--}880^{\circ}\text{C}$. The maximum distance between the two curves is about 3°C . In order to test the whole number relations more accurately, the points must be plotted on a very much larger scale, and fig. 5 shows the results in the neighbourhood of 944°C . the sloping lines showing the mean slopes of the equivalent composition curves. The value of 945.8°C for the silver-cadmium alloy No. 1 is too low by about $0.4\text{--}0.5^{\circ}\text{C}$. to lie on any probable curve through the remaining points for the silver-cadmium alloys. The point for the silver-cadmium alloy No. 3 (Eq. C. 8-15) with 0.4°C . recalescence is 0.5°C . lower than that for alloy No. 2 of practically the same composition but with no recalescence.

* For abbreviation we shall use the term "equivalent composition curve" to describe the liquidus curve of a system when plotted in terms of equivalent compositions. We shall also refer to an alloy of equivalent composition X as "an alloy of Eq. C. X ."

The supercooling tests suggest that not more than 0.2°C . of this difference can be ascribed to the supercooling so that the true liquidus points for alloys Nos. 2 and 3 are probably between the limits 943.9 and 944.2°C ., and the

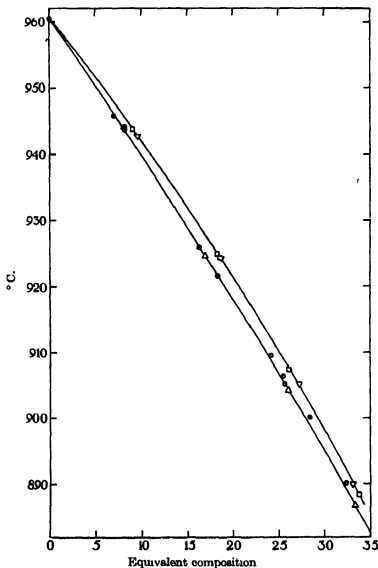


FIG. 4—The liquidus points of the different alloys are plotted against the equivalent compositions with the following notation: AgCd \circ , AgIn \triangle , AgSn \square , AgSb ∇ .

most probable value is between 943.9 and 944.0°C . The four points for the silver-indium alloy No. 1 (Eq. C. 8.02) show that in the absence of recalescence the true liquidus point lies between the limits 944.1 and 944.3°C .

Since the difference of 0.14% in the equivalent composition corresponds with 0.3° C., the equivalent composition curves of the silver-cadmium and silver-indium alloys coincide here to a high degree of accuracy. The three points for the silver-tin alloys show that the true liquidus point for the silver-tin alloy No. 2 (Eq. C. 9.02) lies between 943.7 and 944.0° C. The

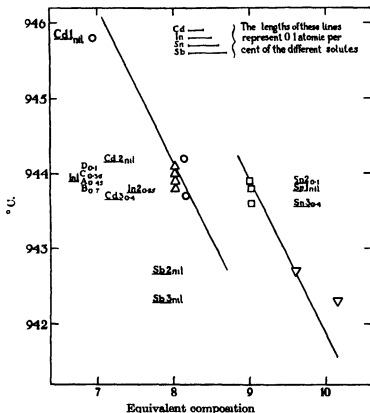


FIG. 5—The liquidus points of the different alloys are plotted against the equivalent compositions with the following notation: AgCd \circ , AgIn Δ , AgSn \square , AgSb ∇ . The subscript figures indicate the recalescence (if any) observed on the cooling curves. The sloping lines show the mean slopes of the equivalent composition curves in this region.

point for the silver-antimony alloy No. 2 (Eq. C. 9.59 by synthesis) lies on the same equivalent composition curve, whilst that for alloy No. 3 (Eq. C. 10.13 by analysis of sample) is about 0.3° C. higher, the difference being within the limits of the sampling method and the analysis.

Fig. 6 shows the points in the region of 925° C. The cooling curves for the

silver-cadmium alloys Nos. 4 and 5 showed 0.2 and 0.15°C . recalescence respectively, and the true liquidus points are probably not more than 0.1 – 0.2°C . higher than those given. The cooling curve for the silver-indium alloy No. 3 showed only 0.05°C . recalescence, so that the true point is probably not more than 0.1°C . higher than that observed since a very slow

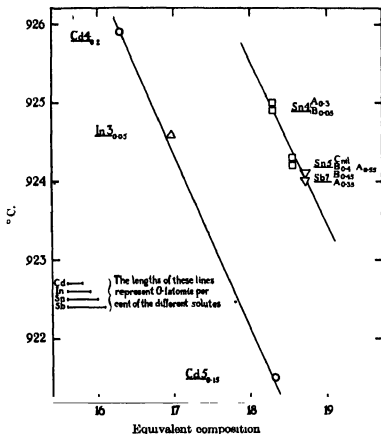


FIG. 6.—The liquidus points for the different alloys are plotted against the equivalent compositions with the following notation: AgCd ○, AgIn △, AgSn □, AgSb ▽. The sloping lines give the mean slope of the equivalent composition curves in this region.

rate of cooling ($0.2^\circ\text{C}/\text{min.}$) was used. When allowance is made for these slight corrections, the points for the silver-cadmium and silver-indium alloys lie on the same equivalent composition curve to within 0.1°C ., so that the agreement with a whole number relation is very precise. The points for the silver-tin alloys Nos. 4 and 5 lie on the same equivalent composition curve as that for the silver-antimony alloy No. 7 to within the limits of the methods

used, but the exact accuracy is less easy to estimate. Reference to Table IV shows that the thermocouple error may account for the fact that the point for the silver-tin alloy No. 4a with 0.3°C . recalescence is 0.1°C . higher than that for alloy No. 4b with practically no recalescence. The repeated experiments for alloys Nos. 5a, 5b, and 5c, for which the thermocouple was steady, show that at the slow rates of cooling employed (1.0 , 0.5 , and $0.5^{\circ}\text{C./min.}$ respectively) the supercooling has little effect, and the point for the silver-antimony alloy No. 7 (rate of cooling $0.5^{\circ}\text{C./min.}$) can hardly be wrong by more than 0.3°C .

Reference to fig. 4 shows that for the points in the neighbourhood of 905 and 890°C . the equivalent composition curves for the silver-tin and silver-antimony alloys still coincide to a high degree of accuracy, and no difference can be detected which is outside the limits of experimental error. For the silver-cadmium and silver-indium alloys the position is less certain. The point for the silver-cadmium alloy No. 8 agrees very closely with the silver-indium equivalent composition curve, but the remaining silver-cadmium points are higher by amounts of the order 1°C . This difference is in the direction to be expected from the errors of the sampling method, and it is significant that it was at this point on the curve that the extracted samples became of uneven composition along their lengths. As these effects were only discovered in the course of the work, this point was only tested for a few of the silver-cadmium alloys. For the silver-cadmium alloys Nos. 1 and 5 with liquidus points at 945.8 and 921.5°C . respectively, the samples were cut in half, and the percentages of cadmium in the two portions differed only by 0.01% . For the silver-cadmium alloys Nos. 6, 7, and 8 with liquidus points at 909.5 , 906.3 , and 905.1°C . respectively, the two halves of the samples differed in composition by 0.14 , 0.07 , and 0.06% respectively, so that errors arising from segregation in the sample during its removal from the melt may here be appreciable.

From the above examination it may be claimed that the equivalent composition curves of the silver-tin and silver-antimony alloys coincide to a very high degree of accuracy over the whole range 960.5 – 880°C . The equivalent composition curves for the silver-cadmium and silver-indium alloys coincide to a very high degree of accuracy over the range 960.5 – 920°C ., and in the range 920 – 880°C . they differ by not more than 1.1°C ., the difference being in the direction to be expected from the experimental errors. The equivalent composition curve for the silver-tin and silver-antimony alloys is, however, definitely higher than that for the silver-cadmium and silver-indium alloys, the maximum difference in the range considered being about 3°C . The initial depressions of the freezing-point of silver produced by

1 atomic % of the above solutes are, therefore, not proportional to the factors 2:3:4:5, but rather to factors of the form $2a:3a:4b:5b$, where b is very slightly less than a , although the whole numbers involved appear exact to within the limits of the experimental methods. It has already been shown (Hume-Rothery, Lewin and Reynolds 1936) that the lattice distortions produced by 1 atomic % of cadmium, indium, tin, and antimony in solid solution in silver vary as the whole numbers 2:3:4:6 to a high degree of accuracy, and it is interesting to note that cadmium and indium have the same, and tin nearly the same relative effects on the depression of freezing-point. Since cadmium and indium have, respectively, one and two more valency electrons than the atom of silver, it is possible that the slight break in the freezing-point relations between indium and tin reveals a discontinuity occurring when the number of valency electrons of the solute atom exceeds two more than that of the solvent.

The authors must express their gratitude to Professor F. Soddy, F.R.S., for his kindness in giving laboratory accommodation and many other facilities which have encouraged them during this work. One author (W. H.-R.) must thank the Royal Society for election to a Research Fellowship. Thanks are also due to the Council of the Royal Society, the Department of Scientific and Industrial Research, the Aeronautical Research Committee, the President and Fellows of Magdalen College, Oxford, and the British Non-Ferrous Metals Research Association for generous grants made to the authors individually to assist their research work. Grateful acknowledgement is also made to Mr J. Gallaher and Mr A. R. Powell of Messrs Johnson, Matthey and Co., Ltd., for their skill in connexion with the chemical analysis and the refining of rare metals.

SUMMARY

A method is described for the accurate determination of the freezing-points of alloys by means of cooling curves in which the temperatures are measured by a thermocouple. The possible sources of error are examined in detail, and by carrying out all determinations under identical experimental conditions, temperature measurements have been made which are consistent to a degree of accuracy varying from ± 0.1 to $\pm 0.3^\circ\text{C}$. Methods for determining the exact composition of the melt at the instant of freezing are investigated. The effect of supercooling is studied, and a method is developed to allow for errors caused by volatilization of one constituent.

The methods developed are used to investigate the liquidus curves of

silver rich silver-cadmium, silver-indium, silver-tin, and silver-antimony alloys. If the freezing-points of the different alloys are plotted against the equivalent compositions (i.e. the atomic percentage of solute multiplied by its valency), the points for the silver-tin and silver-antimony alloys lie on a single curve in the range $960.5\text{--}880^\circ\text{C}$. The points for the silver-cadmium and silver-indium alloys in the range $960.5\text{--}920^\circ\text{C}$. lie on a single curve to a very high degree of accuracy, and in the range $920\text{--}880^\circ\text{C}$. they differ by not more than 1.1°C ., the difference being in the direction to be expected from the experimental errors. The equivalent composition liquidus curve for the silver-antimony and silver-tin alloys is, however, definitely higher than that for the silver-indium and silver-cadmium alloys, the maximum difference in the range considered being about 3°C . The initial depressions of the freezing-point of silver produced by 1 atomic % of the above solutes are, therefore, not proportional to the factors $2:3:4:5$, but to factors of the form $2a:3a:4b:5b$, where b is slightly less than a , although the whole numbers involved appear exact.

REFERENCES

- Heycock, C. T. and Neville, F. H. 1897 *Philos. Trans. A*, 189, 25.
Hume-Rothery, W., Lewin, G. F. and Reynolds, P. W. 1936 *Proc. Roy. Soc. A*, 157, 167.
Hume-Rothery, W., Mabbott, G. W. and Channel-Evans, K. M. 1934 *Philos. Trans. A*, 233, 1.
Reynolds, P. W. and Hume-Rothery, W. 1936 *J. Inst. Met.* (Advance copy.)
Roeser, W. F. and Wensel, H. T. 1933 *Bur. Stand. J. Res., Wash.*, 10, 275.
Weibke, F. and Eggors, H. 1935 *Z. anorg. Chem* 222, 145.
-

The Energy Loss of Cosmic Ray Particles in Metal Plates

BY P. M. S. BLACKETT, F.R.S. AND J. G. WILSON

(Received 23 April 1937)

1—THE EXPERIMENTAL METHOD

Measurements have been made of the energy loss of cosmic ray particles in metal plates, making use of a counter controlled cloud chamber in a magnetic field (Blackett 1936). A metal plate was placed across the centre of the chamber and the energy loss of a ray was deduced from the difference of the curvature of a track above and below the plate. Energy loss measurements by this method have been carried out by Anderson and Neddermeyer (1936) up to an energy of about 4×10^8 e-volts and recently by Crussard and Leprince-Ringuet (1937) up to an energy of 1.2×10^9 e-volts.

The curvature measurements were made mainly by means of the optical null method recently described (Blackett 1937*a*) and this proved invaluable. It would have been hard to obtain so high an accuracy by the usual method of measuring coordinates. The curvature corrections to be applied to the measured curvatures were obtained by measurements on tracks in zero magnetic field (Blackett and Brode 1936). Two separate distortion curves were required, one for the top and one for the bottom of the chamber.

A stereoscopic camera placed to one side of the chamber photographed the tracks by means of an aluminized glass mirror placed at 45° to the axis of the magnet. By this method, a strip of the chamber 10 cm. wide by 24 cm. high could be photographed. The actual maximum effective length of the measureable parts of the track on each side of the plate was about 10 cm.

It is well known that the presence of a metal plate across the centre of a chamber appears to cause a considerable increase in the chamber distortions, especially, but not only, in the neighbourhood of the plate. It was found possible to correlate quantitatively these distortions, which have their origin in convection currents in the chamber, with the thermal conditions of the chamber and of its surroundings; the temperatures of the room, of the base of the chamber which is in thermal contact with the magnet, and of the water supply used to cool locally part of the chamber wall, were found to be the most important variables. A policy of careful temperature control was very successful in reducing these distortions. To effect this control, the chamber was enclosed in a heat-insulating box of fibre board, into which the

pole pieces of the magnet protruded. The camera was built into one side of the box and the illumination was admitted through a thick glass window in the other side. It was found necessary to keep the temperature of the enclosure within 1°C . of that of the magnet. This was effected by keeping the room temperature very near that of the magnet. It was found useful to have a very small amount of water cooling at the bottom of the chamber both to keep moisture off the glass walls and also to stabilize thermally the gas inside the chamber. The maximum power used was only 7 kW, this giving a field of 10,000 gauss at the pole separation of 15 cm. Larger powers were not used, as the increased curvature of the tracks were vitiated by increased distortion due to less favourable thermal conditions arising from the warming up of the magnet. So successful was this method of controlling conditions, that the final errors of the energy measurements of a track on each side of the plate were almost as small as had been obtained in the previous work without a plate, with longer tracks and with a rather greater field.

The convection currents, which it is the object of the present improvement of technique to eliminate, are not those arising from the cooling of the gas in the chamber by the expansion, and which have previously been discussed (Blackett 1936), but are due to the gas currents in the chamber in a steady state, and so are not much reduced by photographing very soon after the expansion, although the necessity for this precaution still remains

2—THE RESULTS

2.1. *The Curvature Corrections*

These were obtained from measurements of eighty-two tracks in zero magnetic field. They consisted of groups of tracks taken under exactly the same conditions as the field tracks described below. These no-field tracks were taken during each run, and were interspaced in time with the field tracks. From these no-field tracks were constructed the distortion curves, giving the curvature correction as a function of position in the chamber, for each half of the chamber for each separate run. The mean error of curvature measurements was such as to give a highest detectable energy E_m of about 2×10^{10} e-volts in a field of 10,000 gauss, and a proportionally lower value in lower fields.

2.2—*The Number of Tracks Stopping in 1 cm. of Lead*

To find this number, a preliminary run was taken in a field of 4000 gauss with all the counters above the chamber. Out of 106 tracks hitting the

plate, fourteen stopped in the plate. Thus about 87 % of the particles can penetrate a 1 cm. lead plate, in agreement with absorption experiments. All the rays which stopped had energies less than 2×10^8 e-volts. The actual energy of each of these rays is shown as a dot at the bottom of fig. 4.

The arrangement used in this run, with all the counters above the plate, is certainly the best method to use, since it records all rays, whatever their behaviour in the plate. However, owing to less favourable geometrical conditions, the yield of tracks per photograph is less than when counters above and below the chamber are used. For this reason, the latter arrangement was employed for all the runs described below. The results quoted above and the discussion of § 3.0 show that no serious errors are introduced, despite the fact that the counter system does not record at all particles which stop in the plate, and only with a small probability those which lose a large part of their energy, or are very much deflected.

2.3—*The Energy Loss in Lead*

A run giving ninety measured tracks in a field of 10,000 gauss with a 1 cm. lead plate gave the energy loss of rays with energies from 3×10^8 to 6×10^8 e-volts. There were in addition twenty-nine measured tracks with energies higher than the upper limit of this range. The results of the energy loss measurements for both positive and negative particles are shown in Table I and in fig. 1. The measured value of the relative energy loss R is given by

$$R = (E_1 - E_2)/x\bar{E} = 2(E_1 - E_2)/x(E_1 + E_2), \quad (1)$$

where E_1 and E_2 are the measured energies above and below the plate and x is the thickness of the plate. Each point in fig. 1 corresponds to the measured value of R plotted against the mean value \bar{E} of the measured energies in the top and bottom halves of the chamber.

A further set of tracks in a field of 3300 gauss with a 0.33 cm. lead plate gave the energy loss of rays between 1.2×10^8 and 4.8×10^8 e-volts. These tracks with the thin plate and with the low field are shown as open circles. The high energy tracks of this run were not measured.

The presence of circles corresponding to a negative energy loss is a result mainly of the errors of the curvature measurements. It is possible, however, that among the particles of very low energy, a few retrograde rays may be present, producing apparent negative energy losses. The presence of such retrograde rays will tend to make the mean measured energy loss rather too small.

Since the error of an energy measurement is given by $\delta E = E^2/E_m$,

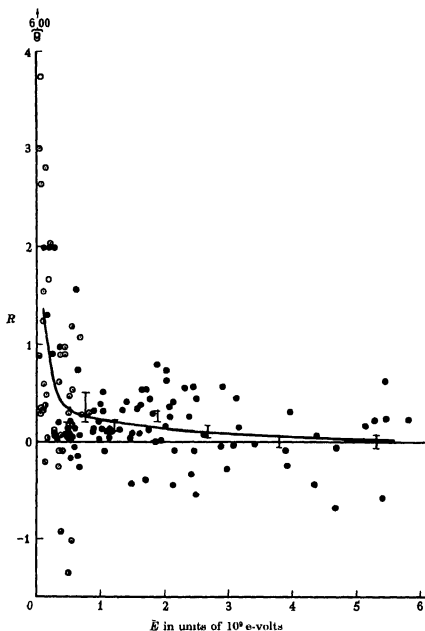


FIG. 1—Plot of energy loss measurements in lead. \circ 0.33 cm. plate, $H = 3300$ gauss; \bullet 1 cm. plate, $H = 10,000$ gauss. Mean curve drawn from equation (19).

where E is the true energy of a track and E_m is the highest detectable energy, the mean error of the relative energy loss R will be given by $\delta R = \sqrt{2} E/E_m$. The factor $\sqrt{2}$ comes from the fact that the energy loss is the difference of two measured energies. This expression only holds for high energies; for low energy tracks the error is rather greater than that given by this formula.

TABLE I—RESULTS OF ENERGY LOSS MEASUREMENTS

	Range of E (10^6 e-volts)	No. of tracks N	Mean energy (10^6 e-volts)	$R = 1/E(dE/dx)$
Lead, 1 cm., 10,000 gauss	0.3-0.6	14	0.45	0.15 ± 0.08
	0.6-1.0	11	0.8	0.38 ± 0.16
	1.0-1.5	17	1.2	0.18 ± 0.04
	1.5-2.25	20	1.9	0.28 ± 0.08
	2.25-3.0	11	2.6	0.12 ± 0.08
	3.0-4.5	9	3.7	0.02 ± 0.06
	4.5-6.0	8	5.3	0.02 ± 0.08
	> 6.0	29	—	—
Lead, 0.33 cm., 3300 gauss	0.03-0.12	12	0.08	1.27 ± 0.25
	0.12-0.30	5	0.13	1.40 ± 0.35
	0.20-0.48	11	0.36	0.20 ± 0.10
Aluminium, 1 cm., 10,000 gauss	< 0.6	4	0.4	0.09 ± 0.02
	0.6-2.0	12	1.2	0.05 ± 0.04
	2.0-4.0	8	3.0	-0.08 ± 0.06

The scatter of the points due to errors of measurement therefore increases with the measured energy, and is about nine times greater for the low-field tracks with the 3.3 mm. plate (open circles, fig. 1) than for the high-field tracks with the 1 cm. plate (black dots).

The open circles in fig. 2 show the mean values of R from Table I plotted against the energy. The vertical lines show the probable errors. Curve 1, fig. 2 is drawn to represent a theoretical expression giving the energy loss (§ 5.2). It represents the observed results satisfactorially.

2.4—The Energy Loss in Aluminium

A short run was made with a 1 cm. aluminium plate in a field of 10,000 gauss. This served as a check on the method of measurement, since a much smaller energy loss was anticipated in aluminium than in lead (§ 6). Actually the energy loss in aluminium was found to be larger than was expected and quite measurable, so that a better check of the technique would have been obtained by the use of a hollow plate of very small absorbing power, which should have led to an experimental distribution of points symmetrical about $R = 0$ for all values of E . Actually the high energy tracks with a 1 cm. lead plate provide a good test of the method of measurement, since they do show

a nearly symmetrical distribution. The black dots in fig. 2 show the results for aluminium. Curves 3 and 4 are two theoretical curves derived on different assumptions to represent the energy loss in aluminium (§ 5.3).

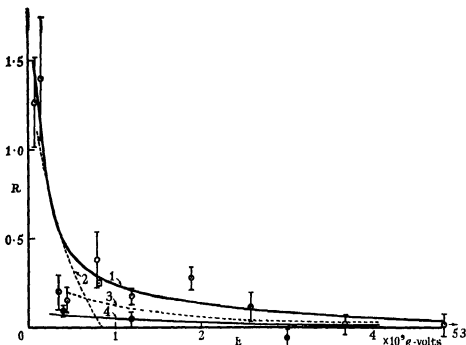


FIG. 2.—Mean relative energy loss of cosmic ray particles in lead and aluminium
 ○ lead: curve 2, equation (9); curve 1, equation (19), $s = 1$. ● aluminium: curve 3, equation (19), $s = 1$; curve 4, equation (19), $s = \frac{1}{2}$. □ lead (Crussard and Leprince-Ringuet 1937).

3—THE DISTRIBUTION OF MEASURED ENERGY LOSSES

The four curves of fig. 3 show the distribution curves of measured energy losses for four ranges of energy:

- (a) $E < 2 \times 10^8$ e-volts; (b) $2 \times 10^8 < E < 8 \times 10^8$ e-volts;
 (c) $8 \times 10^8 < E < 2.2 \times 10^9$ e-volts; (d) $2.2 \times 10^9 < E < 6 \times 10^9$ e-volts.

These curves are the results of two effects, the real energy loss distribution and the effect of errors of measurement. Curve (d), for the highest energy group, is nearly symmetrical about $R = 0$, showing that the mean energy loss is very small, and that the distribution curve is almost entirely due to errors. The rough magnitude of the probable error of the measurements of R is shown

by the horizontal double arrow for each curve, calculated, for high energies, from the expression $\delta R = \sqrt{2E/E_m}$. The smaller the error relative to the breadth of the distribution curve, the more nearly is the true distribution curve correctly represented. The nearly symmetrical form of (d), fig. 3 is an effective guarantee that no large systematic error has been made in the curvature corrections.

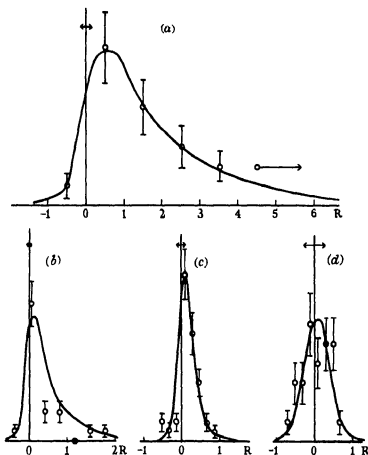


FIG. 3.—Distribution of measured energy losses. (a) $E < 2 \times 10^8$ e.-volts. 3300 gauss, 0.33 cm. lead; 10,000 gauss, 1.0 cm. lead. (b) $2 \times 10^8 < E < 8 \times 10^8$ e.-volts. 10,000 gauss, 1 cm. lead. (c) $8 \times 10^8 < E < 2.2 \times 10^9$ e.-volts. 10,000 gauss, 1 cm. lead. (d) $2.2 \times 10^9 < E < 6 \times 10^9$ e.-volts. 10,000 gauss, 1 cm. lead.

A study of these curves shows clearly that the error involved in using one counter below the chamber is fairly small. For consider a ray with initial energy E_1 and emergent energy E_2 . If $E_2 \leq E_1$, then the mean energy $E \doteq E_1/2$, but the energy loss is $E_1 - E_2 \doteq E_1$, hence the measured value of

$R = (1/E)(\delta E/\delta x) + 2$, when $\delta x = 1$ cm. Hence on curves (b), (c) and (d), fig. 3, a value of R of 2 corresponds to the complete stopping of a particle. For curve (a), which refers to data obtained with a 3.3 mm. plate, a value of R of about 6 corresponds to complete stopping. The form of curves (c) and (d), which approach zero for $R \sim 1$, show that it is most unlikely that there were any appreciable number of particles of these energy ranges which were not recorded because they either stopped or lost a very large part of their energy.

The error may be appreciable for curve (b), and this is a possible cause of the low values of the energy loss found in this region (§ 4). For curve (a) our results are in reasonable agreement with those of Anderson which were not liable to this possible error, so that one can conclude that the error in our case cannot be large, as is also clear from the shape of our curve.

3.1—The Behaviour of Positive and Negative Particles

Table II shows the values of R for positive and negative particles separately, together with the number N of particles measured. For the energy range 3×10^8 to 1.5×10^9 e-volts the values of R are the same within the experimental error. This is not in agreement with the results of Crussard and Leprince-Ringuet, (1937) who find for about the same energy range $R = 0.40$ for negative particles compared with $R = 0.10$ for positives. The mean of these is in agreement with our results. It must be remembered that a relatively small systematic error in the curvature correction curves for the two sides of the plate might produce such an apparent difference.

TABLE II—COMPARISON OF THE ENERGY LOSS OF POSITIVE AND NEGATIVE PARTICLES

Energy (10^8 e-volts)	$N +$	$R +$	$N -$	$R -$
0.3–1.5	26	0.20 ± 0.06	14	0.16 ± 0.04
1.5–6.0	28	0.09 ± 0.10	20	0.23 ± 0.10

For the energy group from 1.5×10^8 to 6×10^8 e-volts, the negative particles show a larger loss, but not much outside the probable error. From fig. 3(d) it can be seen that the observed difference, $R(-) - R(+)$ = 0.14, is only a small fraction of the total width of the distribution curve. It can therefore be concluded that no certain differences between the behaviour of positive and negative particles has yet been shown to exist, though it is quite possible that some difference may exist in some energy range.

3.2—*On the Magnitude of the Systematic Error due to the existence of Random Errors*

Owing to the random error of curvature measurement in the two halves of the chamber, the energy loss measurements may be systematically in error. This follows from the fact that the effect of a random curvature error on the track in the top of the chamber, where the energy is high, is greater than the effect of the same curvature error on the track in the lower part of the chamber, where the energy is lower. The order of magnitude of this systematic error can be calculated as follows, assuming that the random errors have a Gaussian distribution. This has been verified for the no-field tracks by Blackett (1937*a*), and is now verified roughly for the field tracks, as shown by the results of fig. 3 (*d*).

Let the specific curvature spectrum of the incident rays be $N(\sigma)$. Then the probability that a track with a measured curvature σ_0 has a true curvature between σ and $\sigma + d\sigma$ is proportional to

$$N(\sigma) e^{-h^2(\sigma - \sigma_0)^2} d\sigma,$$

and consequently the most probable value E_p of the true energy $E = 1/\sigma$ of the particle is

$$E_p = \frac{\int_{-\infty}^{+\infty} (1/\sigma) N(\sigma) e^{-h^2(\sigma - \sigma_0)^2} d\sigma}{\int_{-\infty}^{+\infty} N(\sigma) e^{-h^2(\sigma - \sigma_0)^2} d\sigma}.$$

There are formal difficulties in the evaluation of the integral in the neighbourhood of $\sigma = 0$, where in addition $N(\sigma)$ is quite unknown, but when $\sigma_0 \gg (1/h)$ ($h = 0.477 E_m$, where E_m is the maximum detectable energy) the important part of the integral does not include this region. This condition is satisfied for the present measurements. It may be shown that the most probable energy E_p is given in terms of the measured energy E_0 by the relation

$$E_p = E_0 [1 + \alpha(E_0/E_m)^2 + \beta(E_0/E_m)^4 + \dots],$$

where α and β are constants depending upon the form of the incident curvature spectrum. The correcting terms are largest for high energies, where the curvature spectrum (Blackett 1937*a*) is nearly constant (this result is also conformed by the present series), and so the correcting terms are readily calculated. Using the probable curvature error as calculated from the

no-field tracks, the following values of the correcting terms have been calculated:

E_0	σ_0	$1 + \alpha(E_0/E_m)^3 + \beta(E_0/E_m)^4 + \dots$
7×10^8 e-volts	0.15	1.02
3×10^9 e-volts	0.3	1.005

For larger curvatures the correction will be still smaller. Thus the correcting factor has a negligible effect on the energy values in the present work. The effect of this correction upon the relative energy loss follows immediately and is roughly the square of the above correcting factor for the corresponding energy, and is thus also negligible.

4—DISCUSSION OF RESULTS

The open circles in fig. 2 show the values of the mean relative energy loss R in lead from Table I plotted against E . Fig. 4 shows the same results but plotted on a larger scale of E , to show up better the results for low energies. The square point is the mean of the values for both positive and negative particles found by Crussard and Leprince-Ringuet (1937), and is seen to lie close to the mean curves.

Anderson and Neddermeyer (1936) have measured the energy loss of rays up to 4×10^9 e-volts. From their results the values of R have been calculated from equation (1) and are shown in Table III and in fig. 4. Anderson and Neddermeyer only measured a ray if it was accompanied by another ray passing through the lower counter; in this way they avoided bias against high energy losses.

TABLE III—ENERGY LOSS MEASUREMENTS OF ANDERSON AND NEDDERMEYER

Range of E (10^8 e-volts)	< 50	50–100	100–150	150–200	200–400
$\bar{E}' = \bar{E} - \Delta E/2$	22	58	93	139	209
R { (Pike's Peak)	1.75	1.37	1.06	1.32	1.30
(Pasadena)	1.43	1.50	1.31	1.53	—
\bar{R}	1.64	1.44	1.63	1.42	1.30

It will be noticed that our two points at $E \doteq 4 \times 10^8$ e-volts both lie well below the mean curve 3, fig. 4, while Anderson's highest energy point for $\bar{E} \doteq 2 \times 10^9$ e-volts lies well above it. Between these two energy values we found very few tracks to measure; there appeared, in fact, to be a minimum in the energy spectrum at this point. While this may well be statistical, it is not impossible that the energy loss may be high in this region, so that, with

the counter arrangement used by us, the particles would not be recorded. The detailed nature of the curve in this region must await further investiga-

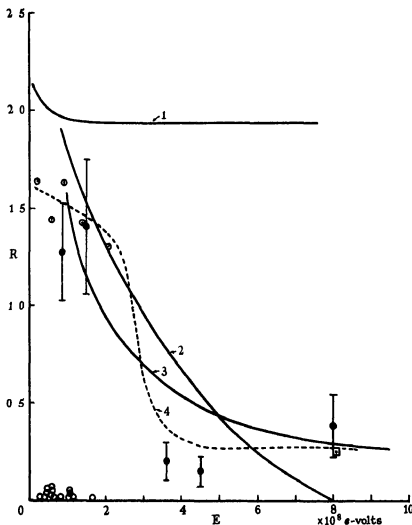


FIG. 4—Relative energy loss of cosmic ray particles in lead. Low energy region. Curve 1, equation (2); curve (2), equation (9); curve (3), equation (19), $s = 1$; curve 4, possible form of energy loss curve. \odot Anderson and Neddemeyer; \square Crussard and Leprince-Ringuet; \bullet Blackett and Wilson; \circ fourteen particles stopping in plate (1 cm. lead, 4000 gauss) out of 120 tracks taken with all counters above the chamber.

tion, in particular, the very low values of R for $E \div 4 \times 10^8$ e-volts require checking, with all the counters above the chamber.

Apart from such possibilities of local anomalies, the main run of the relative energy loss curve seems well established. R has a fairly constant

value of about 1.5 for energies from 0.5 to 2×10^8 e-volts, and then falls rather rapidly to a value of about 0.5 for $E = 4 \times 10^8$ e-volts and then more slowly to values of less than 0.1 for energies of 4×10^9 e-volts and over. The dotted curve 4 is a possible form for the true variation of R with E .

5—COMPARISON WITH THEORY

5.1—The Radiation Loss According to Quantum Mechanics

Bethe and Heitler (1934) have calculated the energy loss by the emission of radiation, during collisions of electrons of very high energy with atoms of atomic number Z , and have obtained the following expression for the cross-section for total stopping of an electron,

$$\sigma = \sigma_0(4 \log 183Z^{-1} + 2/9), \quad (2)$$

$$\text{where} \quad \sigma_0 = Z^2 r_0^2 \alpha, \quad (2a)$$

$$\text{where} \quad r_0 = e^2/mc^2 \text{ and } \alpha = 1/137.$$

$$\text{Then } R \text{ is given by} \quad R = E^{-1}(dE/dx) = \sigma N, \quad (3)$$

where N is the number of atoms per unit area of the absorber.

It is seen that σ is independent of E . Actually, to compare (2) with the observed values, the ionization loss must be added. Curve (1), fig. 4, gives this theoretical energy loss, calculated from (2) and (3) taking into account in addition the ionization loss. It will be seen that the observed values of R are comparable with the theoretical values (actually $R_{\text{obs}}/R_{\text{cal}} \doteq 0.8$) for $E < 2 \times 10^8$ e-volts, but for higher energies the relative energy loss falls sharply to much smaller values.

Since all the particles with $E < 6 \times 10^8$ e-volts are recognizable as electrons from their ionization, there can be no doubt as to the nature of the particles in this energy region. Further, the measurements at higher energies ($E \sim 1.5 \times 10^9$ e-volts) cannot be reconciled with the theoretical value by postulating a large proton component for which the radiation energy loss is negligible, for this is incompatible with the observed distribution of energy losses (fig. 2 (c)).

5.2—Attempted Modifications of the Radiation Formula

Since it has long been clear that the radiation formula (2) must break down at high energies, though the exact value of the energy at which the break-down started was not known, attempts have been made by Williams (1934), by Oppenheimer (1935) and by Nordheim (1936) to modify the formula so

as to reduce the calculated energy loss. No rigid way of doing this is, of course, possible, so the attempts have been based on the semi-classical method of "impact parameters" developed by Williams (1934, 1935) and Weizsäcker (1934). This approximate method gives for the cross-section for total stopping

$$\sigma \sim 4\sigma_0 \log(p_2/p_1), \quad (4)$$

where p_2 and p_1 are the upper and lower values of the impact parameter for which the radiation loss occurs. If p_1 is put equal to $\lambda_0 = \hbar/mc = a_0/\alpha$, where $a_0 = \hbar^2/mc^2$ and equals the radius of the hydrogen atom, and if p_2 is made equal to the screening radius p_F of the Thomas-Fermi statistical atom, which is given by

$$p_F = a_0 Z^{-1} = \lambda_0 Z^{-1} \alpha^{-1}, \quad (5)$$

then

$$\sigma \sim 4\sigma_0 \log(137Z^{-1}), \quad (6)$$

in approximate agreement with the exact expression (2). Williams and Oppenheimer suggested both possible alternative values for p_1 but they did not lead to expressions for σ at all in agreement with experiment. Nordheim, later, has suggested a value for p_1 , which is much more promising. From considerations concerning the difficulties in the simultaneous application of the principles of quantum mechanics and of relativity, it was suggested that for field strengths above the critical value

$$X_c = mc^2/e\lambda_0, \quad (7)$$

there should be no radiation loss. It follows that no radiation loss should occur for impact parameters $p < p_c$, where

$$p_c = \lambda_0 (Z\alpha\xi)^{\frac{1}{2}}, \quad (8)$$

where ξ is the energy of the electron in units of mc^2 . If E is the energy in units of 10^6 e-volts, then $\xi \doteq E/2$.

Inserting this value p_1 in (4), and putting as before $p_2 = p_F$, Nordheim obtained the following cross section

$$\sigma \sim \frac{4}{3}\sigma_0 \log(\xi_c/\xi), \quad (9)$$

where

$$\xi_c = (137)^3 Z^{-3/2}. \quad (10)$$

According to (9), σ decreases as ξ increases, and vanishes when the energy ξ reaches the critical value ξ_c , which for lead equals 830×10^6 e-volts. Curve 2, fig. 2, and curve 2, fig. 4, show the form of (9), when the multiplying factor has been chosen so as to fit the observations at $E = 2 \times 10^6$ e-volts. It will be seen that the observed initial rapid drop is well represented, but that the expression clearly fails for $\xi \sim \xi_c$. It can easily be seen that this behaviour is

due to the inadequate representation of the screening effect of the atomic electrons, obtained by putting p_s equal to the screening radius p_F . The cross section σ vanishes for $\xi = \xi_c$, because p_c from (8) then equals p_F from (5). But it is pointed out by Nordheim that electrons with impact parameters greater than p_F will certainly radiate appreciably. It is thus necessary to introduce the screening in a more consistent manner, for instance by making use of the actual field of the Thomas-Fermi statistical atom.

5.3—Approximate treatment of Screening by the method of Impact parameters

Fermi's expression (1928) for the potential v in an atom is

$$v = (Ze/r) \phi(r/\mu),$$

where $\mu = 0.855a_0 Z^{-1/3} \doteq p_F$ of equation (5). Thus $Z\phi(r/\mu)$ can be considered to be the effective atomic number Z_{eff} at any distance from the nucleus. As the radiation loss depends upon Z^2 we require $\phi^2(x)$ where $x = r/\mu$. Curve 1, fig. 5, shows Fermi's values of $Z_{\text{eff}}/Z = \phi^2(x)$. The approximation made in the impact parameter method by putting the upper integration limit equal to p_F , is equivalent to considering $Z'_{\text{eff}} = Z$ for $x < \mu$, and then zero afterwards (curve 3, fig. 1).

Instead of using the actual Fermi potential it is convenient to make a simple analytic approximation to it. A convenient form to assume is

$$\phi^2(x) = \frac{a^2}{a^2 + x^2}. \quad (11)$$

The dotted curve 2, fig. 5, shows this function, with $a^2 = 0.22$ to give a fit at $x = 1$, where $\phi^2(x) = 0.18$.

For $x \ll a$, when we have $\phi^2(x) = 1$, the lower limit of integration is still given by (8). For $x \gg a$, condition (7) of section 5.2 gives instead approximately

$$p_c^2 = (a\xi Z\alpha)^{2/3} \lambda_0^{4/3}. \quad (12)$$

We will make the simplifying assumption that, for a collision with impact parameter p , the atom can be considered as having an effective atomic number $Z(p) = Z \frac{a^2}{a^2 + x^2}$, where $x = p/\mu$. Following the procedure of Williams (1935 b) we suppose the electron to be at rest and the atom to be moving past it at a distance p with a velocity $v \doteq c$. Then the spectral

distribution which represents the field of the atom moving past the electron with all values of p greater than p_m is

$$I(\nu, p_m) = \frac{4e^2 Z^2}{c} \int_{p_m}^{p_\nu} \frac{\phi^2(p)}{p} dp,$$

where

$$p_\nu = \frac{c\xi}{2\pi\nu}.$$

Inserting $\phi^2(x)$ from (11), and noting that $p_\nu \gg p_m$, we get

$$I(\nu, p_m) = \frac{2e^2 Z^2}{c} \log \left(1 + g \frac{a^2}{p_m^2} \right), \quad (13)$$

where g is a numerical factor differing little from unity.

If we now follow in detail the procedure of Williams (1935 *b*) we get for the absorption cross section

$$\sigma \sim \sigma_0 \log \left(1 + g \frac{a^2}{p_m^2} \right),$$

where σ_0 is given by (2a).

Since $a^2 = 0.22\mu^2 = 0.885^2 \times 0.22 p_\nu^2 = 0.17 p_\nu^2$, and using $p_m = p_c$ from (12), we get finally

$$\sigma \sim \sigma_0 \log \left\{ 1 + g \left(\frac{\xi_c}{\xi} \right)^4 \right\}, \quad (14)$$

where

$$\xi_c = 0.17(137)^2 Z^{-5/3}. \quad (15)$$

This holds for large values of ξ , that is when $p_c \gg a$. For low values of ξ , p_c is given by (8) instead of by (12), and in this case

$$\sigma \sim \sigma_0 \log \left(1 + g \frac{\xi}{\xi_c} \right), \quad (16)$$

when ξ_c is again given by (15).

For high energies, $\xi \gg \xi_c$, either (14) or (16) represents the experimental results almost equally well. For lower energies, $\xi \sim \xi_c$, (16) gives a steeper rise with decreasing ξ and so a better fit.

It is interesting to note that the form of (14) is not much dependent on the form assumed for $\phi^2(x)$ in (11). If $\phi^2(x)$ is taken as proportional to $1/(b+x^n)$ instead of the form of (11), then we find that

$$\sigma \sim \sigma_0 \log \left\{ 1 + g \left(\frac{\xi_c}{\xi} \right)^n \right\}, \quad (17)$$

where

$$\xi_c = 0.17^{2/n} 137^2 Z^{-5/3}, \quad \text{and} \quad s = 2n/(4+n). \quad (18)$$

With $n = 2$, (17) and (18) give (14) and (16).

Now the potential in the outside of the atom, $x \gg 1$, falls off faster than we have assumed by adopting the analytic form (11), in fact $n = 4$ is a more appropriate value than $n = 2$. Using this value in (17) we get again (16) but with a value of ξ_c about 2.4 times larger.

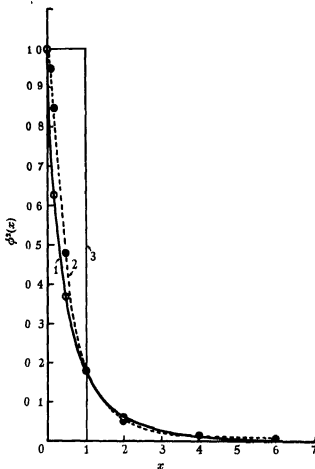


FIG. 5—1, full curve, Fermi $\phi^2(x)$. 2, dotted curve, $\phi^2(x) = \frac{\alpha^2}{\alpha^2 + x^2}$; $\alpha^2 = 0.220$. 3, $Z_{\text{eff}} = Z$, $x < 1$; $Z_{\text{eff}} = 0$, $x > 1$; $x \neq 1$ corresponds to the screening radius p_F , equation (5).

Taking these various considerations into account, we conclude that the relative energy loss is likely to be given by an expression of the form

$$R = AN\sigma_0 \log \left(1 + \left(\frac{E_c}{E} \right)^4 \right), \quad (19)$$

where

$$E_c \approx \epsilon(137)^3 Z^{-5/3} \times 10^6 \text{ e-volts}, \quad (20)$$

and s has a value between $2/3$ and 1 , and ϵ is a constant with the value about 0.1 .

Since the frequency of the emitted radiation does not appear in the logarithmic term of (13), the number of photons emitted with frequency between ν and $\nu + d\nu$ is simply proportional to $d\nu/\nu$. This should represent the distribution of energy losses as observed in fig. 3

6—COMPARISON OF THE MODIFIED NORDHEIM THEORY WITH EXPERIMENT

From (20) for lead, $E_c = 165 \times 10^8$ e-volts and curve 1, fig. 2, and curve 3, fig. 4 show the form of (19), with $s = 1$ when A is chosen to fit the observations in the middle of the spectrum ($R = 0.23$ when $E = 10^9$ e-volts, whence $A \approx 12$). Except at low energies, the experimental results are well represented. But for $E \sim 2 \times 10^8$ e-volts the experimental curve seems to be steeper, being possibly of the form of curve 4, fig. 4

For aluminium, (20) gives $E_c = 3.5 \times 10^9$ e-volts, and with the same value of A and s , curve 3, fig. 2 gives the predicted values. These are higher than the observed points (black dots). Thus though (19) with $s = 1$, represents fairly well the energy loss for lead over a range of energy of nearly 60 to 1 (10^8 e-volts to 6×10^9 e-volts), it gives too high a value for the ratio of the loss in aluminium to that in lead. However, if s is put equal to $2/3$, the ratio of the loss in aluminium to that in lead is given correctly, as can be seen from curve 4, fig. 2, which is calculated from (19) using the value of A as determined for lead.

Considering the very approximate and arbitrary nature of Nordheim's method of limiting the radiation process, the agreement of the observations with an expression of the form (19) is as good as could be expected.

It is interesting to note that the measured energy loss in aluminium is about one-fifth of that in lead for electrons of 10^9 e-volts, although the product NZ^2 for aluminium is only $1/22$ times that for lead, the unexpectedly high energy loss in aluminium is due to the larger values of the logarithmic term resulting from the much higher values of the critical energy.

To get a more exact comparison with the observations, it will be necessary to refine the theoretical calculations in several ways. Particularly for light elements it may be necessary to use Hartree's self-consistent field instead of the Thomas-Fermi statistical field. Then again, the effect of screening must be taken into account correctly, that is, by using a spherically symmetrical atomic model, instead of the pseudo-cylindrical model assumed here as a simplifying assumption (§ 5.3). But such refinements are clearly premature

till the Nordheim condition can be made more precise, and till the arbitrary sharp cut-off at a given value of the impact parameter can be replaced by some theoretically deduced continuous function.

Since Nordheim's condition (7) limits the radiation loss to collisions in which the impact parameter is greater than p_c , as given by (8), and since p_c for high energies is of the order of the screening radius of the atom, the radiation loss is exclusively determined by the outer parts of the atom. Hence it is not impossible that the radiation loss by fast particles in light elements may show periodic properties connected with the different electron shells. It is possible that the fairly sudden drop in energy loss of cosmic ray particles in air, for $E \div 2.5 \times 10^8$ e-volts, as deduced from the energy and absorption spectra (Blackett 1937 *a* and *b*) may be explained in this way, as being connected with a rather sudden falling off of the molecular field in the outer parts of the air molecules. It will be seen that equation (17) gives a more rapid falling off of the radiation loss for high values of s . Of course the approximation (11) to the atomic field is quite inadequate for large z , giving much too large values of Z_{eff} .

An attempt will be made in another paper to give a possible explanation of the conception of a critical field, equation (7), and to avoid the obvious difficulty that this equation is not invariant for a Lorentz transformation

We wish to express our gratitude to Mr A. H. Chapman for his work of taking all the photographs. One of us (J. G. W.) is indebted to the Department of Scientific and Industrial research for a grant to enable this work to be carried out. The cost of the magnet and the apparatus was provided by grants from the Mond Fund of the Royal Society and from the Government Grant Committee.

SUMMARY

1—The energy loss of cosmic ray particles has been measured by measuring their change of curvature in passing through a metal plate placed in the centre of a counter-controlled cloud chamber in a magnetic field. Measurements were made with lead plates 1.0 and 0.33 cm. thick and with an aluminium plate 1.0 cm. thick. The chamber distortion curves giving the curvature corrections as functions of the positions of the tracks in the chamber were obtained by measurement of tracks in zero magnetic field.

2—The results of the energy loss measurements are shown in Table I and in figs. 1, 2 and 4. For lead, the mean relative energy loss

$$R = (1/E) dE/dx$$

is nearly constant and of the order of 1.5 for energies up to about 2×10^8 e-volts, and then falls rather sharply to a value of about 0.4 for $E = 5 \times 10^8$ e-volts and then slowly to a value of the order of 0.03 for $E = 4 \times 10^9$ e-volts.

For aluminium the relative energy loss seems to be about one-fifth of that of lead over the range of energies from 5×10^8 to 2×10^9 e-volts. No certain difference between the behaviour of positive and negative particles has been established.

3—These results, which are shown to be roughly consistent with the previous results of Anderson and Neddermeyer up to energies of 4×10^8 e-volts and those of Crussard and Leprince-Ringuet, are in striking contrast with the predictions of quantum mechanics (curve 1, fig. 4) which predicts a constant value of R about 1.9 for all but the lowest energies (equation 2). There is no possibility of avoiding this discrepancy by assuming protons in the beam, as all the particles with energy less than 6×10^8 e-volts are recognizable as electrons.

A discussion of the ways that have been suggested to modify the radiation formula to give a lower energy loss shows that Nordheim's suggestion, that radiation is not emitted when the field strength exceeds a critical value, leads to fair agreement with the facts for low energies (equation (9)). Its failure for high energies is shown to be due to inadequate representation of the atomic screening. When this is carried out in a more exact manner, by making use of a simple analytic approximation to the Thomas-Fermi field of an atom, an equation (19) is obtained which represents the energy loss results fairly satisfactorily over a range of energy of nearly 60 to 1, within the rather large experimental errors. The following tables give the calculated and observed relative energy loss for lead and for aluminium:

Lead

$E (10^8 \text{ e-volts})$	100	200	400	600	1000	2000	4000
R_{calc} (equation (19) with $s = 1$)	1.46	0.90	0.52	0.37	0.23	0.13	0.06
R_{obs}	1.5	1.4?	0.3?	0.3	0.23	0.13	0.03

Aluminium

$E (10^8 \text{ e-volts})$	500	1000	2000	4000
R_{calc} equation (19) with $s = \frac{1}{2}$	0.063	0.049	0.036	0.026
R_{obs}	0.08	0.05	0.03	?

The distribution of measured energy losses are in reasonable agreement with the predictions of quantum mechanics for the low energies for which this theory holds (fig. 3 (a)).

REFERENCES

- Anderson and Noddermeyer 1936 *Phys. Rev.* **50**, 283.
Bethe and Heitler 1934 *Proc. Roy. Soc. A*, **146**, 83.
Blackett 1936 *Proc. Roy. Soc. A*, **154**, 564
— 1937 *a* *Proc. Roy. Soc. A*, **159**, 1.
— 1937 *b* *Proc. Roy. Soc. A*, **159**, 19.
Blackett and Brode 1936 *Proc. Roy. Soc. A*, **154**, 573.
Crussard and Leprince-Ringuet 1937 *C.R. Acad. Sci., Paris*, **204**, 240
Fermi 1928 *Z. Phys.* **48**, 73.
Nordheim 1936 *Phys. Rev.* **49**, 189.
Oppenheimer 1935 *Phys. Rev.* **47**, 44.
Weizsäcker 1934 *Z. Phys.* **88**, 612.
Williams 1934 *Phys. Rev.* **45**, 729
— 1935 *a* *Nature, Lond.*, **135**, 66.
— 1935 *b* *K. Danske Vidensk. Selsk. (Math. fys.)*, **13**, No. 4.
— 1936 *c* "Kernphysik." Berlin: Springer.
-

On the Relation between Direct and Inverse Methods in Statistics

BY HAROLD JEFFREYS, F.R.S.

(Received 29 January 1937)

1—In my "Scientific Inference" (1931) I gave a discussion of the posterior probability of a parameter based on a series of observations derived from the normal law of errors, when the parameter and its standard error are originally unknown. I afterwards (1932) generalized the result to the case of several unknowns. Dr J. Wishart pointed out to me that the form that I obtained by using the principle of inverse probability is identical with one obtained by direct methods by "Student" (1908). A formula identical with my general one was also given by T. E. Sterne (1934). The direct methods, however, deal with a different problem from the inverse one. They give results about the probability of the observations, or certain functions of them, the true values and the standard error being taken as known, the practical problem is usually to estimate the true value, the observations being known, and this is the problem treated by the inverse method. As the data and the propositions whose probabilities on the data are to be assessed are interchanged in the two cases it appeared to me that the identity of the results in form must be accidental. It turns out, however, that there is a definite reason why they should be identical, and that this throws a light on the use of direct methods for estimation and on their relation to the theory of probability.

Suppose that the true value and the standard error are x and σ ; the observed values are n in number with mean \bar{x} and standard deviation σ' . Then my result (1931, p. 69) was, for previous knowledge k expressing the truth of the normal law but nothing about x and σ ,

$$P(dx | \bar{x}, \sigma', k) \propto \{1 + (x - \bar{x})^2 / \sigma'^2\}^{-1/2} dx; \quad (1)$$

incidentally in the course of the work it turned out that no functions of the observed values other than \bar{x} and σ' are relevant to the result. Now if we put $x - \bar{x} = z\sigma'$ this becomes

$$P(dz | \bar{x}, \sigma', k) \propto (1 + z^2)^{-1/2} dz. \quad (2)$$

But the right side depends only on z , so that in this form \bar{x} and σ' are irrelevant. Hence we can suppress them and write simply

$$P(dz | k) \propto (1 + z^2)^{-1/2} dz. \quad (3)$$

"Student's" method is to find the probability that \bar{x} and σ' will lie in particular ranges, given x and σ ; this is

$$P(d\bar{x}d\sigma' | x, \sigma, k) \propto \sigma'^{n-2} \exp\{-n\sigma'^2/2\sigma^2\} \exp\{-n(x-\bar{x})^2/2\sigma'^2\} d\bar{x}d\sigma'. \quad (4)$$

This breaks up into factors depending on \bar{x} and σ' separately; his proof of this factorization is incomplete, but has been completed more recently, in one way by R. A. Fisher, and in another by J. Wishart and M. S. Bartlett (1933). Now if we write $x - \bar{x} = z\sigma'$ as before, $d\bar{x}d\sigma'$ is transformed to $\sigma' dz d\sigma'$; and then the probability that z is in a particular range dz is

$$\begin{aligned} P(dz | x, \sigma, k) &\propto dz \int_0^\infty \sigma'^{n-1} \exp\left\{-\frac{n\sigma'^2}{2\sigma^2}(1+z^2)\right\} d\sigma' \\ &\propto (1+z^2)^{-1/2} dz, \end{aligned} \quad (5)$$

on suppressing factors independent of z . This is independent of x and σ ; hence we can suppress them and write

$$P(dz | k) \propto (1+z^2)^{-1/2} dz, \quad (6)$$

which is "Student's" result, and is identical in form with (3).

Both (3) and (6) have the property that, subject in the one case to our only data beyond k being x and σ , and in the other to the only data beyond k being \bar{x} and σ' , the distribution of the probability of z depends on z alone. Consequently they are cases of the kind where Bernoulli's theorem is applicable. We can integrate and get results of the form

$$P(\sigma'z_1 < x - \bar{x} < \sigma'z_2 | k) = F(z_2) - F(z_1). \quad (7)$$

If then we consider only cases where this difference has a particular value, we shall expect by Bernoulli's theorem that out of a large number of trials the fraction such that $x - \bar{x}$ will lie within the range $\sigma'z_1$ to $\sigma'z_2$ will approach $F(z_2) - F(z_1)$. In consequence, therefore, of the irrelevance to z of x and σ in the one case, \bar{x} and σ' in the other, the two methods make identical predictions in the long run. Further, the claim made for the direct method that the assertions based on its results will be correct in a given fraction of trials in the long run is justified.

It is important to notice that this application of Bernoulli's theorem, like all others, depends on a certain condition, that the probability of the event at every trial is independent of the results of previous trials. If we acquire information at one trial that is relevant to the results expected from later trials the theorem is untrue. Thus the possibility of applying the theorem depends entirely on the supposition that other experience gives no information relevant to z in a new experiment.* This is true for (6); the only data

* A very full discussion of Bernoulli's theorem is given by J. M. Keynes (1921, Chapter 29).

relevant to \bar{x} and σ' before the experiment are the truth of the normal law of error and the values of x and σ for that particular type of observation.

In practice we do not often want the probability of \bar{x} and σ' for known x and σ ; we usually want that of x and σ for known \bar{x} and σ' . In other words we want $P(dx d\sigma | \bar{x}, \sigma', k)$, or, if we are not directly interested in σ , we want $P(dz | \bar{x}, \sigma', k)$. But if this is independent of \bar{x} and σ' it must be equal to $P(dz | k)$, which is already known from (6); and the statement that z is in a particular range dz is equivalent to

$$\sigma' z \leq x - \bar{x} \leq \sigma'(z + dz), \quad (8)$$

which, when σ' , z and dz are specified, may equally well be regarded as a proposition about \bar{x} knowing x or about x knowing \bar{x} . Thus provided that \bar{x} and σ' are irrelevant to z we can infer (3) from (6). "Student's" result can therefore in these conditions be used to determine a posterior probability, which is what we usually want. This argument, it will be noticed, does not involve Bernoulli's theorem, but it does involve the notion of irrelevance, without which Bernoulli's theorem is inapplicable. It is not obvious that \bar{x} and σ' are irrelevant to z ; in fact if we have any knowledge of them, whether (3) is true or not, (5) is certainly false. k cannot include any information about the observations. To take some extreme cases, if both x and \bar{x} are known, the probability on the data that $x - \bar{x}$ lies in any particular range is not given by (5), but is certainty for some ranges and impossibility for others. Similarly if σ' is known the factor depending on it in (4) is zero for all ranges $d\sigma'$ except one including the actual value. Integration with regard to σ' then gives the same factor for all values of x , and the probability distribution for \bar{x} is $\propto \exp\{-n(x - \bar{x})^2/2\sigma^2\} d\bar{x}$, from which σ' has disappeared, and which is quite different from (6). Thus while the probability distribution for z is given by (6) provided we know nothing more than x and σ , and possibly less, it is quite different if we have any other information about \bar{x} and σ' . Of course when actual measurement has given both \bar{x} and σ' , z is definitely known and its probability is certainty for one range and impossibility for all others.

These considerations for extreme cases are entirely obvious, but serve to illustrate the unsatisfactoriness of estimates of the probability of a proposition by itself without explicit statement of the data. We can assert a frequency in the long run from (7) whether x and σ are known in advance for each experiment or not; they may vary or be always the same; but if we know either of them we cannot assert the same probability distribution for z in any one experiment for which we also know anything about \bar{x} or σ' . For some values of \bar{x} and σ' the probability of a given z will be less, for others

more, than is given by (6). In such a case we should naturally use our additional information (which need not be an exact evaluation*) to obtain a different formula, which we should expect to lead more often to correct inferences in the long run. The suppression of x and σ to derive (6) from (5), and the hypothesis that \bar{x} and σ' can then be added to k without altering the results, are therefore essential to the argument.

The absence of information relevant to \bar{x} and σ' , apart from x , σ , and the truth of the normal law, being essential to (5), it is clear that the argument cannot hold if at any time information about both pairs of quantities appears simultaneously in the data. But we have got to (2) from (6) by incorporating \bar{x} and σ' in the data and supposing that they make no difference; hence the argument can be valid only if nothing is known about x and σ . These are the conditions contemplated in my proof of (2). The same thing can be seen in another way. Referring to (8) we see that the distribution of the probability of x given \bar{x} and σ' has the property that if another series of observations gives a different value of \bar{x} , we shall displace the maximum of the distribution for x by just the difference between the two values of \bar{x} ; and if another series gives a different value of σ' we shall multiply the scale of the distribution for x by the ratio of the two values of σ' . But if x was already known to lie within a particular range the probability that x was in that range would be unity whatever \bar{x} was, and we could not displace the distribution bodily with \bar{x} . Similarly if known sources of error say that σ cannot be less than a certain amount, we shall be reluctant to apply the proportionality rule if σ' turns out to be less than that amount, and shall interpret the close agreement as accidental.

Thus the use of "Student's" rule to infer anything about x given \bar{x} and σ' involves an unstated assumption, that $P(dz | \bar{x}, \sigma', k)$ is independent of \bar{x} and σ' , and this appears to be equivalent to saying that we know nothing initially about x and σ ; when we make this assumption my result follows and the direct and inverse methods agree.

1.1—Further, it is possible, if we assume that the result is correct for two observations, to work back from it to the distribution of prior probability that I used to express previous ignorance with respect to x and σ . The formula is symmetrical about \bar{x} ; but if the prior probability of x was not uniformly distributed this would not hold for all values of \bar{x} . If the prior probability density for x varied in a range of order σ' about \bar{x} , the posterior probability density would not have a maximum at \bar{x} . The necessary con-

* For instance, \bar{x} and σ' , instead of being calculated, might be estimated from the median and the average residual without regard to sign.

dition that the maximum shall be at \bar{x} for all values of \bar{x} and σ' is therefore that the prior probability density for x is uniform. We suppose that of σ independent of that of x , so that we can write

$$P(dx d\sigma | k) \propto f(\sigma) dx d\sigma,$$

where $f(\sigma)$ is to be found. Now if we simply take the case $n = 2$, the probability that z will be between ± 1 is $\frac{1}{2}$, by (7), a result noticed by "Student". This is equivalent to saying that there is an even chance that the true value will be between the first two observed values, given these values, whatever they may be. It has been proved (Jeffreys 1936, p. 425) that if the law of error is such that the true value is the median of the probability distribution for the observed values, given the true value and the law of error, a sufficient condition for this to hold is $f(\sigma) \propto 1/\sigma$. We have to find out whether this condition is necessary. Evidently the result cannot hold if $f(\sigma)$ involves any parameter other than σ , for if it did this would appear explicitly in the result. It can therefore be only a power of σ (sin σ and log σ , for instance, have no numerical values if σ is a length). It is convenient at this stage to use the precision constant h instead of the standard error and to write

$$f(\sigma) d\sigma = g(h) dh/h. \quad (9)$$

Then if the first two observations are $\pm a$,

$$P(dx dh | \pm a, k) \propto h g(h) \exp\{-2h^2(x^2 + a^2)\} dx dh. \quad (10)$$

The probabilities that x is between $\pm a$, and $\pm \infty$, are then respectively proportional to I_1 and I_2 , where

$$I_1 = \int_0^\infty g(h) \exp(-2h^2 a^2) \operatorname{erf}(\sqrt{2} ha) dh, \quad (11)$$

$$I_2 = \int_0^\infty g(h) \exp(-2h^2 a^2) dh. \quad (12)$$

Thus the result required is that $I_1 = \frac{1}{2} I_2$. The condition for it to hold is that

$$\int_0^\infty g(h) \exp(-2h^2 a^2) (2 \operatorname{erf} \sqrt{2} ha - 1) dh = 0 \quad (13)$$

for all a . This is easily verified for $g(h)$ constant, and the integrand changes sign once within the range of integration. It cannot be true if $g(h)$ is any monotonic function such as a power, since the integrand for a positive power would be multiplied by a larger factor when it is positive than when it is negative, and conversely. Thus $g(h) = \text{constant}$ is the only possible solution, and we have $f(\sigma) \propto 1/\sigma$, which is my previous assessment of the prior

probability for the case where σ is unknown. Further, the formula (1) has been proved in detail from this assessment for any value of n .

Thus we have the following results. "Student's" formula as it stands is an assessment of the probability of the ratio of the error of the mean to the standard deviation, given at most the true value and the standard error, and no other information relevant to the mean and standard deviation. It is not directly applicable as it stands to the assessment of the probability of the true value given the mean and the standard deviation, but can be made applicable by means of a further hypothesis, namely that the distribution of the probability of z , given the mean and the standard deviation, is the same for all values of these quantities. This appears to be equivalent to the supposition that the mean and standard error are previously unknown, and the only distribution of prior probability that leads to it is the one that I have previously used to express the condition that they are unknown. Thus the hypothesis that "Student's" formula is applicable to the probability of the true value given the observations is completely equivalent to my assessment of the prior probability and corresponds to the same conditions for its applicability.

1.2—An objection that has been made to the law

$$P(d\sigma | k) \propto d\sigma / \sigma$$

is that on integration from 0 to $+\infty$ it diverges at both limits. I have pointed out (Jeffreys 1933, p. 530) that without such divergence it would express a degree of knowledge inconsistent with what the law is designed to say, namely, that we know nothing about the value of σ . It appears, however, that some confusion has arisen from the first axiom and a convention in the theory of probability. The first axiom, in its present form, is (Jeffreys 1931, p. 11):

If we have two sets of data p and p' , and two propositions q and q' , and we consider the probabilities of q given p and of q' given p' , then whatever p , p' , q , q' may be, the probability of q given p is either greater than, equal to, or less than that of q' given p' .

This is a postulate about probability as an elementary idea, and comes before the association of probabilities with numbers is mentioned. It simply says that probabilities can be ordered. It is, however, unnecessarily stringent. We cannot decide between courses of action on different data; if two sets of data are relevant and known, we make our decision on data pp' . If only p is known, we can decide only on data p . Further, it turns out that the postulate that we can compare probabilities on different data is never used.

It appears to be used in the product formula, which is fundamental,

$$P(pq | r) = P(p | r) P(q | pr).$$

Two factors involving different data appear on the right. But the arguments for this formula rest on a convention, that the probability number unity is to be associated with certainty on whatever data. If we do not compare probabilities on different data there is no objection to using different numbers to express certainty on different data, though of course we must always use the same number on the same data. Now consider the argument in § 2.32 of my book, but suppose that on data h certainty is denoted by a , and that on data qh certainty is denoted by b . Then (1) of that section becomes

$$P(q | h) = am/n,$$

since we are retaining the addition rule; and (2) is

$$P(pq | h) = al/n = am/n.l/m.$$

But

$$P(p | qh) = bl/m;$$

whence

$$P(pq | h) = P(q | h) P(p | qh) / P(q | qh),$$

the certainty in the denominator removing the factor b . The formula thus expresses the equality of the ratios of two pairs of probabilities, one pair on data h , the other on data qh . With this modification we can therefore proceed without needing to compare probabilities on different data. Bayes's proof from expectation can be adapted similarly

The principle of inverse probability as it stands contains only ratios of probabilities on the same data. The modified product formula contains an extra factor, but as this will appear in all estimates on the same data it will cancel, and there is no change. Further, all estimates of probabilities are either statements of previous ignorance or built up from them by means of this principle as new data are taken into account, and are therefore comparable from the start, being calculated values. Thus the comparability of probabilities on different data, if we choose always to attach the same number to certainty, is proved in the natural course of events as occasion arises, and needs no special axiom. But as it is never used it is not of the slightest importance whether it is proved or not.

It is in fact convenient in most cases to denote certainty by 1, but we are at liberty to use any other number if it is convenient in any particular case. Now a parameter capable of an infinite range of values is liable to be such a case; for zero is in any case associated with impossibility, and if we take the probability that the parameter is in an infinite range as unity the probability that it is in any finite one will be zero unless it is concentrated about

some value in such a way as to give a convergent integral. In that case the distribution would say that it is already practically certain that the true value is within a certain finite stretch, and would therefore be unsuited to express previous ignorance. Thus when we wish to express previous ignorance over an infinite range we should expect to find it convenient to represent certainty by infinity. But in that case the divergence of the integral is not only expected but demanded. The objection to the divergence of the integral is simply due to the use of a convention in a case where it is unsuitable. On the other hand, in the problem of measuring one unknown, as soon as we have two observations there is material to form an estimate of the standard error, and the integrals of the posterior probability converge; thus we can again take unity to express certainty. The complication arises only for the case of previous ignorance over an infinite range.

There is a similar difficulty with regard to x if this also is unrestricted apart from the observations. No objection has been made, as far as I know, to the fact that $\int P(dx|k)$ diverges; but the same complication is present and can be dealt with in the same way.

I have pointed out already that convergence can be saved if x and σ are in fact restricted initially by any vague knowledge that imposes upper and lower limits to them, and that provided these limits are sufficiently widely separated this restriction has a negligible effect on the results. This is the usual practical case, but in the very beginnings of a subject there appears to be no such restriction, and the above modification is needed if we are to avoid dealing with ratios of infinitesimals.

The $d\sigma/\sigma$ law appears to have a wide application, and is not restricted to standard errors. It is applicable whenever there is no reason to prefer a direct estimate of a quantity to one of some power. I first gave it on the ground that for the normal law of error there is no reason to prefer either h or σ to the other as an estimate; but as $h\sigma$ is constant the distributions dh/h and $d\sigma/\sigma$ are equivalent. Similarly in a determination of a density, if we say that the density is unknown we say also that the specific volume is unknown, but the distributions dp/p and dv/v are equivalent, and no other pair of the same form would be. Eddington has pointed out that some methods of determining the charge on an electron give e , others e^2 , again, since de/e is proportional to de^2/e^2 , the same law will do for either. The rule appears to apply to any one-signed quantity that may have any value between 0 and ∞ . Even for an elementary case like the volume of a cubical vessel, we can measure it by filling it with water and transferring the water to a measuring glass, or we can measure the side and cube it. The distributions

dv/v and dl/l are equivalent if $v = l^2$, but this applies to no other pair of distributions of the same form. This result is somewhat surprising, since the prior probability of a length would usually be taken as uniformly distributed. But uniform distribution would say that if we suggest any finite value for l , the true value is certain to exceed it. It must be remarked that the only test of a probability distribution used to express ignorance is that it shall not make a positive statement of a kind that we could not possibly make without previous knowledge.

It is somewhat remarkable that, though the work of Bayes and Laplace has been attacked to the extent of rejecting the principle of inverse probability altogether, this modification of their assessment of prior probability was not mentioned until my "Scientific Inference". The fact that a uniform distribution could never apply both to a quantity and to any power of it had often been mentioned, but the only modification considered was to reject the prior probability altogether; the obvious fact that complete equivalence between powers could be achieved by taking a logarithmic distribution seems somehow to have escaped notice.

Has the uniform distribution any applications, then? They can occur only when there is no question of comparing powers. Thus for sampling, we want to estimate r , the number of members of the parent class with a given property. Distributions for r and r^2 in this case cannot be expected to have the same form, since r admits all possible integral values, up to the whole number of the class, and r^2 only some. In this problem there is no objection to the uniform distribution, apart from the extreme values.

Another case where the uniform distribution is obviously applicable is the longitude of a newly discovered asteroid. In all such cases the possible values of the parameter are limited in both directions, and this initial limitation appears necessary to the uniform distribution.

In practice most physical parameters are statements of differences. Even the length of an object is usually the difference between the observed positions of its ends against a scale. But when we assert a significant difference we are not making a *a priori* statement. If we have only two measures we have no means of saying without other evidence whether the difference between them is wholly systematic, wholly random, or a mixture of the two. It might appear that the length of a body and the standard error of a measurement are two independent unknowns, and that their joint prior probability is proportional to $dl d\sigma/\sigma$. This *a priori* hypothesis of independence is, however, not justified. It may be true, but we are not entitled to assume it without positive evidence. It is a statement of causality, and can be made only if experimental evidence shows that l and σ have no significant corre-

lation. One may indeed suspect that there is a correlation, since large quantities are usually measured with a larger absolute uncertainty than small ones, and a length less than σ/\sqrt{n} cannot be asserted to be different from zero from observational evidence. The separation of l and σ in this way therefore presupposes that l has already been shown to be different from zero, and this can be done only by showing that it is more probable that the differences between measures on the two ends of the body contain a systematic part than that they are wholly due to random variation. The primary unknown, then, is neither l nor σ , but the expectation of the separation of two measures on different ends of the body. If we call this s , we have $s^2 = l^2 + \sigma^2$, and $P(ds|h) \propto ds/s$. But now l is limited to the range between $\pm s$, and the natural form for the prior probability is $P(dl|sh) = dl/2s$. This is found (Jeffreys 1936, pp. 432-7) to lead to a satisfactory test for the proposition that l is different from zero, and also to a complete determination of the posterior probability distribution for l .

If u and v are two quantities known to be positive, but otherwise unrestricted, uv and v/u have the same properties, and their prior probabilities would be expected to follow the same law if this is general. To evaluate them it is necessary first to suppose some limits imposed on u and v , thus

$$a < u < A, \quad b < v < B,$$

and we can take $b/a < B/A$. Then if $v/u = \phi$,

$$P(d\phi|h) = \iint P(du dv|h) \propto \iint du dv/uv$$

taken over the region such that $u\phi < v < u(\phi + d\phi)$. The results are

	$P(d\phi h)$
$\phi < \frac{b}{A}$	0
$\frac{b}{A} < \phi < \frac{b}{a}$	$\frac{d\phi}{\phi} \log \frac{A\phi}{b}$
$\frac{b}{a} < \phi < \frac{B}{A}$	$\frac{d\phi}{\phi} \log \frac{A}{a}$
$\frac{B}{A} < \phi < \frac{B}{a}$	$\frac{d\phi}{\phi} \log \frac{B}{a\phi}$
$\frac{B}{a} < \phi$	0

So long as A and B are very large compared with a and b the second factor varies very slowly in comparison with $1/\phi$, and in the limit its variation is so

slow that it can be neglected. The disturbance of the $d\phi/\phi$ rule is serious only near the extremes and is introduced only by the mathematical device adopted to ensure convergence. Thus the rule applies to the ratio v/u . Further it applies to $1/u$ and therefore to uv . It is therefore consistent.

1.3—The uniform distribution for the prior probability of x or l is equivalent to the adoption of the mean as the best value, if the prior probability is not uniform the maximum posterior probability is not at the mean. The mean is sometimes advocated on the ground that if the normal law of error holds, the mean is unbiased. Now at this point there is liable to be a confusion. It is perfectly true that the mean, given the true value, is as likely to have a positive error as a negative one. But this is not the same as saying that the true value, given the mean, is as likely to be on one side of it as the other. This applies only if we have no previous reason for expecting it to be on one side rather than the other of the mean as found in that particular experiment. Consider, for instance, a large population of quantities whose true values are distributed approximately normally with standard variation s . We measure a sample of them individually, each measure having standard error σ . Any individual measure, given the true value, is as likely as not to have a positive error. But, given the mean, there are more true values within a distance of order σ on the side of it towards the centre of the distribution than on the other side. The true value, given the mean, is therefore more likely to be nearer the centre than farther away. Thus the true distribution is systematically different from the observed one, and in describing it we effectively apply to each measure a correction based on information provided by the others. The validity of an unbiased estimate therefore also depends on particular conditions of previous knowledge. In practice this type of problem would be treated by correcting the observed variation for the known standard error of observation, thus estimating the true distribution as one with a smaller variation than the observed distribution. I think that statisticians generally agree about the legitimacy of such a process, but the systematic changes that it introduces are contrary to the hypothesis that the so-called unbiased estimate is necessarily the best estimate. The additional hypothesis that enters in using the mean in such a case is that the probability distribution of the true value given the mean is the same for all values of the mean; whereas actually it differs according to the position of the mean with respect to the centre of the distribution.

1.4—Since it has turned out that the condition that “Student’s” formula can be interpreted as a posterior probability is that there shall be no previous knowledge of x and σ , it may be suggested that this is in general the con-

dition that direct methods can be used to assess a posterior probability. It is well known that if the prior probability of x is uniformly distributed, the posterior probability density of x , when the standard error is already known, is proportional to the likelihood. The same is true for a sampling ratio on the Bayes-Laplace theory. Thus several cases of this type of equivalence are known. It appears that if the distribution involves true values a, b, c, \dots , sufficient statistics for which are $\alpha, \beta, \gamma, \dots$, and if the likelihood involves a and α only through $a - \alpha$, b and β only through $b - \beta$, c and γ only through c/γ , and so on, the condition needed will be that a change in α and β shall be considered best represented by equal changes in a and b , and that a change in γ shall be considered best represented by a proportional change in c . But this is equivalent to taking

$$P(dadbdc | k) \propto da db dc / c$$

This is a common case, but does not appear to be general. When a and b are differences to be inferred from the observations themselves, the possibility that they are zero is usually, perhaps always, taken seriously to begin with, and is only discarded if a significance test shows that the zero value is not supported by the observations. Thus a finite amount of the prior probability is concentrated in the zero value, and the law is not of the type just given. This applies equally whether a and b are contingencies, correlations, or coefficients of assigned functions to be used in approximating to a series of measured values. The possibility that the variation is wholly random has always to be disposed of before the question of assigning values other than zero to the differences arises. This problem has received no unique solution by direct methods, but can be solved by the use of the principle of inverse probability.

The problem of the volume of the cubical vessel, on this theory, would be analysed as follows. The fundamental quantity for the length measurement is $s^2 = l^2 + \sigma^2$; given s , half the prior probability of l is uniformly distributed from $-s$ to $+s$. The other half is concentrated at $l = 0$. The first problem then is to find whether the mean found for the length differs significantly from zero. If it does we have the posterior probability distribution for the true value. Similarly we can find the distribution for the probability of the volume. If necessary we can test whether the values are consistent with $v = l^3$, if they are, the results can be combined and improved values for both l and v will be obtained. Any relation between two physical magnitudes appears to rest on an analysis of this kind.

It appears therefore that results obtained by direct methods cannot in general be interpreted as posterior probabilities, though there are cases

where they can be used as an intermediate stage in determining the latter. If the question asked concerns the true values given the observed values, the answer must depend on the nature of any previous restrictions on the true values that may exist, this must be expressed through the prior probability, and the inverse theorem must be used. Incidentally this method appears to be easier mathematically in most cases I have met than the direct methods that give superficially similar results, but really answer a different question.

2—*Statistical Mechanics*. At first sight this subject appears to be an application of direct methods of probability theory, but there are several difficulties of interpretation. If a system starts in a given state and the laws of classical dynamics are assumed, the position and velocity of any component (molecule, proton, or electron) at any later instant are theoretically completely determinate. In this case any function of the state at a later instant either has an assigned value or not, and no question of degrees of probability arises. In practice, however, the initial state is known only in certain general features and not in detail. Only a very slight uncertainty in the position or velocity of any one component will affect the identity of the first other component that it strikes, and this will affect the whole subsequent motion. Thus uncertainties of the initial state introduce questions of the probability of later states. Since every component will ultimately affect every other, directly or indirectly, these probabilities must deal with properties of the system as a whole. It may turn out that for almost all initial states these properties, after a specified time, will lie within certain ranges, and in particular that the number of components within any given range of velocity will have approximately a particular value. Such properties will then be physical magnitudes in Campbell's sense, and the relations between them will be physical laws. There will be exceptions if the initial motions are such that no collisions will ever occur, but we are entitled to ignore such a possibility in the absence of positive evidence for it.

In wave mechanics there is an ultimate uncertainty of the initial state, and the last possibility, remote in any case, is excluded. The observational uncertainty of the initial positions and velocities, however, far exceeds the Heisenberg uncertainty, which merely gives a limit to the accuracy that might be attainable but has not been attained.

It does not appear that the theory of statistical mechanics has ever been derived in this way, and the existence of permanent parameters associated with the distribution appears to be taken as an experimental fact and not inferred from the theory. The methods actually used fall into two types, represented by the Boltzmann H theorem and the Gibbs ensemble.

2.1—The H theorem appears to show that a gas started off arbitrarily will approach a Maxwellian distribution of velocity. Unfortunately, however, the proof is defective. It considers collisions in detail; the probabilities of any two molecules lying within assigned ranges of position, and assigned ranges of velocity, are taken to be given separately, and the joint probability that both are in the indicated ranges simultaneously is taken to be the product of the separate probabilities. Now on the face of it this is a most dangerous hypothesis. It amounts to asserting

$$P(pq|r) = P(p|r)P(q|r). \quad (1)$$

The correct probability statement is

$$P(pq|r) = P(p|r)P(q|pr) \quad (2)$$

In words, the common statement, that the probability that two propositions are both true is the product of the probabilities that they are true separately, is not a general rule. The joint probability is the product of the probability of the first and the probability of the second *given the first* and the original data. It is only when the first is irrelevant to the second that (1) is correct. The kind of error that may arise if it is adopted may be illustrated by considering a population in which half the members have blue eyes and half brown eyes. The probability that a new individual will have a blue right eye is $\frac{1}{2}$, the probability that he will have a blue left one is $\frac{1}{2}$; and if we apply (1) we find that the probabilities that he will have two blue eyes, two brown ones, and one blue and one brown are respectively $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$. If on the other hand we use (2), r now including the observed result that the two eyes in known individuals are the same colour, the probability that the left eye of the new individual will be the same colour as his right is 1, and we get the values $\frac{1}{2}$, $\frac{1}{2}$, 0. The form (1) contains no means of expressing the fact that the eyes are of the same colour.

Now the first equation used in the H theorem asserts that if the probability of the centre of a molecule having velocities in particular ranges $du dv dw$ is $f(u, v, w) du dv dw$, the probability of a collision in time dt between two molecules is

$$v^2 f(u, v, w) f(u', v', w') V \cos \theta \cdot \sigma^2 du dv dw du' dv' dw' dt dx dy dz$$

(Jeans 1921, p. 20). Here (xyz) is the position of the centre of the first molecule and v the numerical density. The main criticism of this formula is that the positions and velocities of the two molecules before collision are supposed irrelevant to one another. This may be correct for the final state, but when, for instance, gas is admitted to an evacuated vessel, there are originally variations of mean density and velocity from place to place, and the position

TABLE III.—COMPARISON AND EXTRAPOLATION OF BEATTIE-BRIDGEMAN EQUATION WITH EXPERIMENTAL RESULTS

$T = 25.05^\circ$												
d	18.75	32.7823	69.1538	91.1131	112.181	153.793	226.708	308.608	386.741	491.852	596.455	
p	18.537	20.865	50.489	57.871	62.155	—	—	—	74.739	308.620	1325.17	
pv B.-B.	0.98742	0.90859	0.72638	0.63353	0.57796	—	—	—	1.26675	2.67150	4.89221	
pv M.	0.98866	0.91102	0.73011	0.63516	0.55406	—	—	—	0.19325	0.62747	2.22175	
Δpv	-0.00124	-0.00243	-0.00373	-0.00163	-0.00390	—	—	—	+1.07350	+2.04403	+2.67046	
$T = 40.71^\circ$												
p	20.446	33.475	59.525	70.782	79.158	90.683	103.257	121.776	180.898	518.686	1697.79	
pv B.-B.	1.09956	1.01881	0.85591	0.77336	0.70649	0.61344	0.59348	0.52827	1.35264	2.60267	4.56825	
pv M.	1.09045	1.02115	0.86076	0.77688	0.70563	0.58964	0.45546	0.39460	0.46775	1.05456	2.84847	
Δpv	-0.00089	-0.00234	-0.00485	-0.00350	+0.00086	+0.02360	+0.13802	+0.43367	+0.88489	+1.54811	+1.72178	
$T = 96.77^\circ$												
p	24.248	—	76.745	—	110.947	—	184.991	—	403.563	—	2419.47	
pv B.-B.	1.29469	—	1.10968	—	0.99215	—	0.90579	—	1.50429	—	4.00918	
pv M.	1.29325	—	1.10977	—	0.98900	—	0.81599	—	1.04347	—	4.05726	
Δpv	+0.00144	—	-0.00009	—	+0.00315	—	+0.08980	—	+0.46082	—	-0.04808	
$T = 160.14^\circ$												
p	28.031	—	93.426	—	141.534	—	266.859	—	628.422	—	3117.01	
pv B.-B.	1.49935	—	1.35748	—	1.26652	—	1.19477	—	1.63954	—	3.53457	
pv M.	1.49498	—	1.35099	—	1.26166	—	1.17443	—	1.62492	—	5.22590	
Δpv	+0.00437	—	+0.00649	—	+0.00486	—	+0.02034	—	+0.01462	—	-1.69133	

point. His values differ from the authors' at densities below 300 Å on the average 0.3 %, but at densities round 400 Å about 4 %.

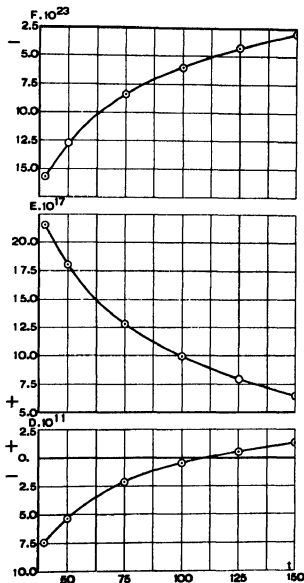


FIG. 3

Beattie and Bridgeman (1928) have published a "new equation of state for fluids" with which we have compared our observations in Table III.

The equation for densities up to 112 fits the experiments with an accuracy of 0.7 %.

The isotherms, published by Nijhoff, Michels and Gerver (1930), up to 1800 atm. give an average deviation of 0.2 % and a maximum of 0.35 %.

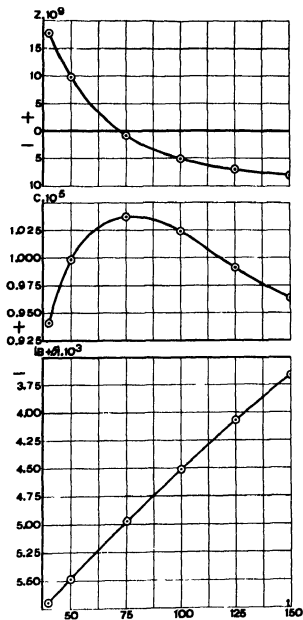


FIG. 2

Keesom (1903) published isotherms in the neighbourhood of the critical

TABLE II— $(pv_{\text{calc}} - pv_{\text{exp}}) \times 10^5$

Gas:	d	0°	25-053°	29-900°	31-037°	32-075°	40-105°	49-712°	75-260°	99-767°	125-007°	150-140°
	18-75	-1	-3	-8	7	2	+42	+37	+20	+23	+32	-5
	23-4609	1	-1	0	0	5	+32	+21	+6	+12	+18	-6
	28-1756	-2	7	8	0	3	+28	+21	+4	+5	+17	-4
	32-7823	4	-1	3	-7	-4	+8	+2	+7	+9	+7	-9
	37-0251	-1	0	-1	2	2	-1	-12	-6	-5	+9	-16
	42-4239	-3	0	0	3	0	-8	-7	-11	-12	+7	-9
	47-1236	1	-2	-4	-2	-8	-28	-17	-10	-8	-3	-8
	51-9162	-	-2	-2	-5	-6	-28	-21	-11	-6	-	-
	45-7631	-	-2	-	1	3	-21	-13	-25	-12	-15	-6
	69-1536	-	5	3	-6	1	-51	-43	-18	-26	-32	-16
	91-1131	-	-3	-2	3	6	-32	-12	-6	-8	-15	+10
	112-181	-	1	1	-2	-5	+8	+12	+10	0	+4	+20
	133-474	-	-	-	-4	1	+38	+34	+23	+19	+14	+28
	163-793	-	-	-	6	0	+74	+42	+23	+25	+11	+30
	170-439	-	-	-	-2	-	+82	+49	+21	+22	+3	+15
	188-541	-	-	-	-	-	-	+33	+21	+22	+4	+22
	206-448	-	-	-	-	-	-	0	+11	+5	-17	-2
	206-465	-	-	-	-	-	+7	-2	-1	+9	+24	-6
	226-708	-	-	-	-	-	-41	-34	-19	+14	+2	-36
	247-483	-	-	-	-	-	-91	-63	-30	-19	-15	-40
	267-766	-	-	-	-	-	-112	-66	-32	-23	-18	-31
	288-060	-	-	-	-	-	-93	-55	-23	-20	-12	-22
	308-608	-	-	-	-	-	-48	+26	-8	-7	+4	+3
	329-365	-	-	-	-	-	+25	+18	+12	+8	+24	+31
	348-296	-	-	-	-	-	+98	+57	+29	+23	+33	+72
	352-168	-	-	-	-	-	+103	+78	+25	-27	-44	-29
	356-741	-	16	-12	-96	-63	+9	+67	+15	+8	-2	+6
	432-259	-	-6	-20	-117	-81	+9	+11	-6	+7	+30	+51
	457-002	-	4	-3	18	19	-124	-93	-	-	-	-99
	491-852	-	-11	13	96	69	-191	-155	-17	+2	-	-
	527-299	-	11	-7	50	29	-73	-54	-29	+7	-35	+39
	553-322	-	-4	7	-111	-80	+136	+101	+50	-18	+40	-3
	596-455	-	1	-3	38	29	+80	-37	-20	+8	-12	-5

TABLE I

t	A	$(B + \beta) \times 10^3$	$C \times 10^4$	$Z \times 10^3$	$D \times 10^{13}$	$E \times 10^{18}$	$F \times 10^{24}$
Gas:							
40-105	1-154650	-5-691941	9-41470	17 75226	-74-64248	215-2911	-157 2770
49-712	1-190062	-5-487750	9 98347	9-76829	-53 50756	180-2757	-127-3692
75-260	1-284231	-4-972395	10-37167	-0-80821	-21 96426	128 0081	-84-1916
99-767	1-374593	-4-514801	10-24786	-5 26154	-5 16369	99-4225	-60-9761
125-007	1-467597	-4-070506	9-91486	-7 05498	5-48658	79-2425	-42-7539
150-140	1-560237	-3 661328	9-64656	-8-00258	13 42336	64-4629	-30-3258
0	1-006834	-6-837046	11 39205	—	-151-4235	—	—
25-053	1-099169	-6-099705	10-69337	14-27129	-170 3612	8913 056	-252457-2
29-900	1-117035	-6-029717	14-63283	-79 33782	618-6174	-15153-460	204050-9
31-037	1-121226	-5-958189	11-60349	-10 20406	60-7083	-965 879	6334-6
32-075	1-125052	-5 928158	11-33589	-6-74054	46-9183	-984-129	7614-8
Liquid:							
25-053	0-523412	-1-129210	2-72154	-13 23712	9-02762	88-36283	-51-10890
29 900	0-218906	1-555123	-4-44812	-9-14362	19-08726	62-26349	-30 95412
31-037	0-239829	1-368326	0-74904	-40-69024	80-10010	-24-54438	40-95583
32-075	0-331725	1-162219	-4-21878	-11-29488	29-06571	37-27420	-5 20579

It is therefore obvious that under these circumstances it will be dangerous to draw conclusions from the value of the coefficients of the series evaluation.

In the series evaluation of his results on nitrogen, Gerver (1934) found that the values for B calculated from the isotherm data in the high pressure range, were not the same as those from the low pressure isotherms of Otto, Michels and Wouters (1934). He ascribed this to an impurity present in the gas used. It now seems more probable however that the divergence is due to the reason given above.

The series evaluation, however, is still of importance for the calculation of p at a given density under the condition that the deviation curve between experimental and calculated values is known. Therefore the series evaluation of type 4,

$$pv = A + (B + \beta)d + Cd^2 + Zd^3 + Dd^4 + Ed^5 + Fd^6,$$

has been carried out for the temperatures between 40 and 150° C.

For the temperatures below 40° C, it was impossible to fit the experiments with one series. Two series were therefore calculated, one for the range of the gas densities and one for the range of the liquid densities. Use was made of the densities of the saturated vapour and of the liquid under vapour pressure (i.e. pv at critical density) as published by Michels, A., Blaisse and Michels, C. (1937) The coefficients for the different series are given in Table I.

The deviations between the experimental values of pv and those calculated from the series evaluation are shown in Table II.

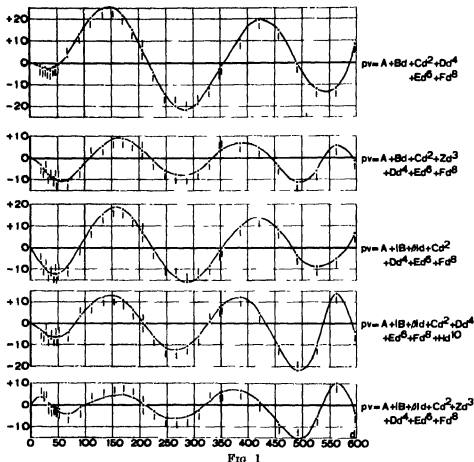
The values for the coefficients of the series between 40 and 150° C. are plotted as functions of temperature in figs. 2 and 3. As can be seen, they all lay on smooth curves. This facilitates interpolation at other temperatures. It may be remarked that in the previous publication, where the results of the series evaluation for the low pressure data were evaluated, only the values of B and not those of the other coefficients fitted on smooth curves (Michels, A. and C. 1935).

The data given above, together with the values for $\left(\frac{dp}{dt}\right)_d$ published previously, allow the calculation of pv at any intermediate temperature and density and they also open the possibility of comparing the isotherm results published with those of other authors.

Calculation showed that the average deviation from Amagat's isotherms up to 1000 atm.* is about $\frac{1}{2}$ %, although occasionally deviations as high as $1\frac{1}{2}$ % occur.

* "International Critical Tables", 3, 11.

the difference between experimental and calculated values has again increased.



The conclusion seems therefore allowed that the isotherm data obtained at the higher densities cannot be expressed with an ordinary power series. The reason can be seen from the values of each of the terms of the series. Point 33 gives for instance:

$pv = 2.84647$	$Zd^3 = 2.07278$
$A = 1.19006$	$Dd^4 = -6.77216$
$(B + \beta)d = -3.27320$	$Ed^6 = 8.11718$
$Cd^2 = 3.55171$	$Fd^8 = -2.04027$

These figures show that for the higher densities the series is not convergent.

later state is a dynamical consequence of the first, this will not in general be true for the second, and the theorem appears to be untrue. Nevertheless if we consider only impulsive encounters and very short time intervals the theorem holds. We can now interchange the momenta of the particles taking part in an encounter; the positions are instantaneously unchanged and can therefore be interchanged, without altering the probabilities; and the positions and momenta of all other particles are instantaneously unaffected by the interchange and can also be interchanged. Then the theorem for the complete system can be applied, and the theorem of detailed balancing is proved for impulsive changes.

2.7—Many of the results of quantum statistics are extensions of the above classical results when treated in terms of the Gibbs ensemble; it seems possible therefore that the above methods can be adapted to quantum theory. One warning may however be needed. The $|\psi^2|$ of wave mechanics is interpreted as a probability, but it is not clear on what data it is expressed. The wave equation replaces the equations of motion, which are exact in classical mechanics; thus the uncertainty of the initial conditions, which appears in the above statement of the classical theory, is fundamentally different from the uncertainty expressed by $|\psi^2|$. In the quantum adaptation it will be replaced by an uncertainty of the initial value of ψ , and the problem will be, as in the classical theory, to show how the uncertainty of the initial conditions can lead to assessments of the probability distribution for certain functions of the variables that after some time are practically independent of the time.

SUMMARY

The probability distribution of functions of the observed values, given the true value and the standard error, is not obviously the same as that of functions of the true value and standard error given the observations. Direct statistical methods appear to depend for their validity on identifying these two distributions. It is shown here that, in a case discussed as a direct problem by "Student" and as an inverse one by the present author, the application of "Student's" result for estimation involves an unstated assumption, which is equivalent to the condition that the true value and standard error are initially unknown, and that if this assumption is valid it is completely equivalent to the author's assessment of the prior probability to express this condition.

The foundations of classical statistical mechanics are discussed with

in which, if μ is the mass of a molecule, the range is such that

$$\frac{1}{2\mu} (p_1^2 + \dots + p_m^2) \leq E_1; \quad \frac{1}{2\mu} (p_{m+1}^2 + \dots + p_n^2) \leq E - E_1; \quad (9)$$

and for given E_1

$$\frac{\partial p_n}{\partial E} = \frac{\partial p_n}{\partial (E - E_1)}. \quad (10)$$

The integral separates into factors and the result is proportional to

$$E_1^{m-1} (E - E_1)^{1/2(n-m)-1} dE_1. \quad (11)$$

Now put

$$E = ns^2/2\mu, \quad m = 1. \quad (12)$$

Then the distribution for one component is easily found to be nearly proportional to

$$\exp(-p_1^2/2s^2) dp_1 \quad (13)$$

for n large. This is Maxwell's law.

For m a moderate fraction of n , the probability of E_1 is concentrated closely about a value such that $E_1/E = m/n$. Thus the rule of partition of energy in approximate proportion to the number of degrees of freedom is established.

2.8—Jeans and Fowler use the H theorem to prove the principle of detailed balancing. Since the present treatment deals with the assembly as a whole, this principle needs restatement. Let us consider the probability that a distribution within a definite small range about p, q at time t will be followed by one about p^*, q^* at time t^* . Since the state at the later time is uniquely specified by the laws of dynamics in terms of that at the earlier, we can take the element $dp_1 \dots dp_n dq_1 \dots dq_n$ to correspond completely to the element $dp_1^* \dots dp_n^* dq_1^* \dots dq_n^*$; and the extensions of the two elements are equal by the theorem of the Jacobian. Then the probability that the system will be in the first state at the time t and in the second state at time t^* is equal to the probability of the first state if the states correspond and otherwise zero. The probability that it will be in the second state at time t and in the first state at time t^* is similarly the probability of the first state at time t^* , and in the steady state this is the same as its probability at time t . Thus the theorem is true for corresponding states of the whole system at different times. The principle of detailed balancing requires the interchange of only part of the variables between the two states. Suppose the variables at time t grouped into two sets w_1 and w_2 , and those at time t^* into sets w_1^* and w_2^* , we have to consider the probabilities that the system will be in state w_1 and w_2 at time t , w_1^* and w_2^* at time t^* , and that it will be in state w_1^* and w_2 at time t , and w_1 and w_2^* at time t^* . If for the first proposition the

constant; hence dE disappears. The variables retained are now restricted to such values that the contribution to the known value of E left to be supplied by p_n must be positive. The rule of equipartition follows at once. Suppose for instance that the p 's are Cartesian components of momentum. The probability that any particular one, p_1 say, is in a particular range is the integral of g with respect to all combinations of the variables, subject to the contribution to the energy from p_1 having a value in the given range. But this has the same form whichever element we take as p_1 , and all have therefore the same probability distribution. The distributions are not, however, quite independent. It is formally possible for one component to absorb the whole of the energy; but then the others are restricted to be zero, and the distributions cannot be combined according to (1). The postulate of independence is therefore not exact even in the steady state.

The uniform distribution found for g when the energy is given is equivalent to the uniform distribution of weights adopted by Fowler; his "number of weighted complexions" corresponding to a given state is then seen to be proportional to the probability of the state, the other factor being his total number of complexions. His results can therefore be interpreted as genuine probabilities for the individual assembly with given total energy in the steady state, without reference to the ensemble. Similarly his averages over the ensemble are seen to be identical with expectations for the individual assembly. The weight is mentioned by Fowler (p. 9) as having been identified with *a priori* probability; but this analysis shows that it is not the prior probability of a hypothesis as in the theory of probability, but (apart from a constant factor) the probability of a result given the hypothesis. Any information about the initial state, apart from the first integrals that can be supposed known, turns out to be irrelevant to g , and similar systems started off with the same values of these functions will be expected to settle into indistinguishable final states. Thus results found from the probability distribution apply to a wide range of initial states and are therefore physical laws. It will be noticed that the effect of introducing the ensemble is to hide the progressive distortion of the state-elements for the individual assembly, and it is this distortion that ultimately removes systematic observable effects due to the initial conditions.

Consider now, in the case of a gas of hard spherical molecules, the probability that any m components of momentum carry energy E_1 in a range dE_1 out of the total E . We can use E_1 to eliminate p_m ; the required probability is then proportional to

$$dE_1 \int \dots \int dp_1 \dots dp_{m-1} \frac{\partial p_m}{\partial E_1} dp_{m+1} \dots dp_{n-1} \frac{\partial p_n}{\partial E} \quad (8)$$

original distribution of rotations persists. If experiments can be devised to make this rotation communicate itself to the apparatus it can therefore be investigated quite separately from the translations. This is a case where g breaks up into factors. An enclosure that is a surface of revolution is another case where functions other than the energy remain constant, since collisions with the boundary do not affect the angular momentum about the axis, and g must therefore be a function of both the energy and the angular momentum. It may be doubted however whether such a perfect surface of revolution exists. The usual restrictions on the nature of g therefore arise naturally in this treatment.

At the same time we have an answer to the difficulty about reversibility of path. Given g , the theory of probability will predict the frequency of random variations of given amount. But it will not predict any return to the initial state because all possible initial states make contributions to g , and cannot be disentangled. If a Maxwell demon reversed all the velocities exactly at a given instant, the initial state might be recovered, but actual experimental methods cannot do it. Any error in the projection will affect the whole motion, and the result will again be a distribution specified, as closely as is determinable, by g . If a vessel is divided into two halves by a partition, the two halves are filled with different gases, and the partition is removed, the gases mix in a few minutes. Random motions might conceivably unmix them in a long enough time, but the asymmetry between the times needed to mix and separate the gases depends fundamentally on the fact that the initial departure from uniformity is far beyond the normal range of random fluctuation, and can be expected to exist only if it is introduced by external disturbance. Changes from it are then traceable to correlations between neighbouring molecules—which, for an arbitrary initial distribution of density or velocity, are precisely what the alleged proof of the H theorem neglects.

2.5—The probability that a function of the variables is within a specified range is then to be got by integrating g with regard to all the unknowns, the limits being given by the conditions that the function must lie within this range; if we express certainty by unity the denominator will be the integral over all possible values. If we consider a free enclosure with the total linear and angular momenta zero, g is a function of the energy alone; then we can use the energy to eliminate one of the unknowns, say p_n (at this stage we may suppress accents), and the probability element is

$$g(E) dp_1 \dots dp_{n-1} dq_1 \dots dq_n (\partial p_n / \partial E) dE.$$

If E is known to lie in a given small range, $\int g(E) dE$ through this range is

properties of molecules over a certain range of position and time; but the range of position, after a certain time, is much greater than the coordinate thicknesses of the element, and the range of time is greater than the time taken for a considerable change in the probability that the state is within a given element of the space. Thus the expected result of any observation will involve probabilities integrated over such ranges that f varies considerably. The tendency of the act of observing itself is therefore to smooth out the irregularities of f ; so long as our observations are of this type, the longer the time since the start the more thorough will the averaging be, and the results will be expected to conform to a uniform distribution of probability density. The observed fact that only certain functions of the initial conditions are correlated with measurable properties after a long time thus receives a qualitative explanation. But we notice that some functions of the variables cannot alter with time; such are energy, in general, and linear and angular momentum in a free system. Points with widely different values of these will never lie close together in the illustrative space, and averaging will therefore not remove their differences. Hence the ultimate probability density, which we now see is an approximation, will involve these functions explicitly.

Let us denote this function by $g(p' \dots q' \dots)$. Evidently for gas in a free vessel, or for a star cluster with a large enough gravitational potential to prevent escape, the variables must be such that the energy and the linear and angular momenta remain constant and equal to their initial values; thus the values of $2n - 7$ of the variables specify the remaining 7. For a system in a fixed enclosure we have additional constraints. The motion of the enclosure is given, collisions do not alter the energy, but any systematic velocity will be reversed by collision with the boundary and be averaged out in observing conditions. Thus the expectation of the linear momenta will be zero in the steady state. This can be achieved if g is a function of the energy of the motion relative to the enclosure. For an enclosure rotating with a given angular velocity about a fixed point we can consider the function $F = \Sigma \mu \dot{x}_i^2$, for spherical molecules; here μ is the mass of a molecule and x_i its position relative to axes rotating with the enclosure. It is easily seen that F is not altered by collisions between molecules; for collisions with the boundary the normal velocity relative to the boundary is reversed and the tangential one unaltered, so that F is again unaltered. Hence the required steady motion is given by taking g to be a function of F . It follows at once that in steady motion the mean velocity at any point corresponds to rotation with the enclosure like a rigid body. On the other hand for smooth spherical molecules no interactions can alter the rotation of any molecule, so that the

appears to contradict the fundamental belief of statistical mechanics that the state of a gas becomes steady and that after a certain time a finite number of parameters are enough to specify it.

It appears that this argument is another way of stating the principle of reversibility, which has also been offered as an objection to the H theorem. Given the coordinates and momenta at time t and the equations of motion, it is mathematically possible to evaluate those at any previous time and therefore to find out what previous time corresponds to the initial conditions. It would take an assembly of the order of n Maxwell demons to find the coordinates and momenta and to do the calculation, but mathematics is not concerned with that.

2.4—Some progress can be made if we regard the variables as plotted in space of $2n$ dimensions. If the initial state is exactly specified the subsequent states lie on a curve. If we consider the set of states corresponding to small deviations about this initial state, they will correspond to an elementary volume in the space, which moves with time, but in such a way as to keep a constant volume dv , and the probability that the system at time t will be within this volume will always be $f dv$. But the shape of the element will alter greatly. If it is originally rectangular the difference in the values of p_1 on two sides will imply a steady difference of \dot{q}_1 ; thus after a certain time two sides originally with p_1 and q_1 constant will not be perpendicular. The element will therefore lengthen in some directions in the space and shorten in others so as to keep the volume constant. If a collision occurs the momenta are all discontinuously changed, the coordinate volume and momentum volume are separately unaltered, but the path is displaced bodily. But whatever the element does the probability density within it remains the same. Now it is fundamental that the initial conditions are not specified with such precision that we can predict all collisions, or even one. Different elements corresponding to moderate values of f and slightly different initial conditions will be displaced at different times in all sorts of ways; but they will all have the property that they become steadily longer in some directions and thinner in others, which will always be inclined to the axes. In any direction parallel to an axis they will therefore in general become thinner. The result will be that regions with different values of f will become closer together in any direction along an axis. If we persist in considering a rectangular element at time t and treating f as constant within it, the element must become smaller and smaller and ultimately fall below molecular dimensions. Thus the results will become meaningless unless we can observe the individual molecule. What we actually can observe are average

But on substituting for p'_1 and q'_1 from Hamilton's equations we see that the coefficient of dt is zero. Hence the Jacobian in (4) is independent of the time and must be permanently unity. This still applies if collisions are taken into account, this can be proved by direct solution of the collision problem for two bodies, or we can regard the collisions as the limit of a continuous movement. If when two bodies come into contact, or when one of them collides with the boundary, a conservative repulsive force comes into play in proportion to the amount of the further approach, we have a Hamiltonian system with formally continuous movement, the limit of which when the repulsive force increases sufficiently fast will be the result of an impulsive collision. This result of course replaces Liouville's theorem.

Now suppose that the probability distribution for $p' \dots q' \dots$ is of the form (3) except that it may involve t explicitly. Then by comparison with (4) we have

$$P(dp'_1 \dots dp'_n dq'_1 \dots dq'_n | k) = g(p'_1 \dots p'_n q'_1 \dots q'_n, t) dp'_1 \dots dq'_1 \dots \quad (6)$$

$$= f(p_1 \dots p_n q_1 \dots q_n) dp'_1 \dots dq'_1 \dots \quad (7)$$

The distribution at any time other than zero is therefore to be got by substituting for the initial values in terms of those at time t . Now it may turn out that g does not involve t explicitly. But if it does not it must be identical in form with f , since we can make t tend to zero, and all the variables to their initial values. But by definition $g(p' \dots q' \dots t)$ is equal to $f(p \dots q \dots)$ for all t . Hence if g does not involve t explicitly, f is unaltered when the variables p and q are replaced by the values at any other time consistent with the equations of motion and the initial conditions. In other words $f(p' \dots q' \dots) = \text{constant}$ must be a first integral of the equations of motion, not involving the time. Now in general there are only a finite number of such integrals, which may be used to eliminate an equal number of variables; if this does not lead to exact determinations of all the other variables f does involve functions of the variables other than the first integrals. Hence the function g must involve t explicitly unless the initial conditions are so stringent that the whole of the variables are exactly specified when a number of them equal to the number of the first integrals of the equations of motion are given. The latter case is plainly highly exceptional. It appears to follow that the probability distribution with regard to $p' \dots q' \dots$ cannot ordinarily be steady. Any uncertainty of the initial conditions, beyond the first integrals, will affect the probability of a given set of values at any later time in such a way as to make it variable with time. Alternatively, if we use the principle of inverse probability, there will be many functions of the variables whose values will give information about the time elapsed since the start. Now this

are made by Fowler, who, however, adopts the ensemble as a working method while recognizing its limitations.

It appears therefore that the two methods suffer from different defects. The Boltzmann method attempts to test the variation with time, but fails to prove a given type of variation because it does not treat the assembly as a whole. The Gibbs method does treat it as a whole, but in such a way that it cannot reveal any variation with time, and it is uncertain how far the results may be expected to apply to a single assembly. The proposition that an assembly would almost certainly tend to equipartition remains unproved, and the second law of thermodynamics is unexplained.

2.3—An analysis based on the theory of probability appears to be possible. It must be noticed first that any result obtained, if it is to express a physical law, or if the quantities involved in it are to be physical magnitudes in Campbell's sense (1920, Chapters 10-13, also Jeffreys 1931, Chapters 4 and 6), must have a high probability for every assembly separately. It is therefore useless and undesirable to try to combine assemblies. Now suppose that the state of an assembly is specified at time 0 by momenta p_1, \dots, p_n and coordinates q_1, \dots, q_n . Corresponding quantities at time t are indicated by accents. Then on classical dynamics all the coordinates and momenta at time t are determinate functions of those at time 0. There is however an uncertainty of the initial state, so that the actual values of the initial coordinates and momenta have a probability distribution of the form

$$P(dp_1 \dots dp_n dq_1 \dots dq_n | k) = f(p_1 \dots p_n q_1 \dots q_n) dp_1 \dots dp_n dq_1 \dots dq_n. \quad (3')$$

Then the corresponding distribution at time t is

$$P(dp'_1 \dots dp'_n dq'_1 \dots dq'_n | k) \\ = f(p_1 \dots p_n q_1 \dots q_n) \frac{\partial(p_1 \dots p_n q_1 \dots q_n)}{\partial(p'_1 \dots p'_n q'_1 \dots q'_n)} dp'_1 \dots dp'_n dq'_1 \dots dq'_n \quad (4)$$

The Jacobian is identically 1 if the system is Hamiltonian. If we use double accents for quantities at time $t + dt$, we have $p''_1 = p'_1 + p'_1 dt$ and so on, thus

$$\frac{\partial(p''_1 \dots p''_n q''_1 \dots q''_n)}{\partial(p'_1 \dots p'_n q'_1 \dots q'_n)} = \left| \begin{array}{cccc} 1 + \frac{\partial p'_1}{\partial p'_1} dt & \frac{\partial p'_1}{\partial p'_2} dt & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & 1 + \frac{\partial q'_1}{\partial q'_1} dt & \dots \\ \dots & \dots & \dots & \dots \end{array} \right| \\ = 1 + \left(\frac{\partial p'_1}{\partial p'_1} + \dots + \frac{\partial q'_1}{\partial q'_1} + \dots \right) dt + O(dt^2). \quad (5)$$

and velocity of one molecule are relevant to the probability of another being within a specified distance and to that of its velocity being within a specified range. The decomposition of the probability of a collision of given type into factors is therefore not legitimate.

Assuming this form, the argument proceeds to estimate the rate of change in the number of molecules within a specified range of velocity. This vanishes if the distribution is Maxwellian. The function $H = \iiint f \log f \, du \, dv \, dw$ is then constructed, and it is shown that $\partial H / \partial t$ (or rather its expectation) is zero if f is Maxwellian and otherwise negative. It is inferred that if f is originally non-Maxwellian H will decrease, and that f will ultimately approach the Maxwellian form. The proof is however vitiated by the postulate of independence, which is untrue for the stages of the motion that are being considered. (Nor indeed does it appear to be proved that H does not tend to $-\infty$.) Nor does the proof that df/dt is zero imply that f is constant, even if at some stage there is a lack of correlation that will permit the separation into factors, for the velocities before collision, this does not prove that the state is permanent unless it is also shown that the velocities after collision are independent. The argument therefore does not show either that a non-Maxwellian distribution will become Maxwellian or that a Maxwellian one once attained will be permanent.

2.2—In the Gibbs method the single system (the assembly of Fowler) is replaced by an ensemble of many other systems, and the probability that a particular assembly is in a given state, with respect to some property, is replaced by the fraction of the assemblies in the ensemble that have the property. This device has been introduced into pure probability theory also, but it leads only to complications. What happens is that properties are found for the average values of parameters over all assemblies, and that these are then applied to the separate assemblies as if they held for all, but they may not hold even for one. The ensemble cannot be simply a repetition of the individual system, the other systems included in it must differ in some respects. The averaging will then remove the differences and retain the properties expected to be physically significant, but until we have decided which these are we do not know what additional assemblies to construct mentally and combine with the first. If we decide wrongly we may get an average that does not apply to the original assembly or may fail to find one that does. Further, no comparison is made between states of the same system at different times; gas that has just been admitted to a vessel is treated as comparable with gas that has been in it a week, and there is no indication of the direction of any change that may take place. Some of these criticisms

It may be possible to find coefficients which allow representation of the experimental data up to the critical density, with an estimated accuracy of 10 % (Beattie, Hadlock, Poffenberger, 1935); for still higher densities there is little hope of any agreement.

The series evaluation as developed in our paper fits the results better than the equation.

Under those circumstances it does not seem justified to recalculate the coefficients of the equation over a higher density range.

SUMMARY

It is common practice to represent isotherm data by a series expression either with powers of the pressure or of the density, e.g. either

$$pv = A + Bp + Cp^2 + \dots$$

or

$$pv = A + Bd + Cd^2 + \dots$$

Previously it has been shown that for isotherm data above 500 atm. the first series is not adequate.

The present paper shows that also the second series cannot represent accurately the isotherm data of CO₂ up till 3000 atm. as published by the authors.

The series representation is, however, still of value for interpolation purposes, if deviation curves between experimental data and series are known. The isotherm data have therefore been expanded in series of the type

$$pv = A + Bd + Cd^2 + Zd^3 + Dd^4 + Ed^5 + Fd^6.$$

The value of a coefficient depends on the density region for which the series is applied, the number of coefficients, etc. and it is therefore dangerous to draw too rigid physical conclusions from the numerical values

With the help of the series the results published have been compared with those published by other authors.

REFERENCES

- Beattie, J. A. and Bridgeman, O. C. 1928 *Proc. Amer. Acad. Sci.* **63**, 220.
 Beattie, J. A., Hadlock, C. and Poffenberger, N. 1935 *J. Chem. Phys.* **3**, 93.
 Fowler, R. H. 1929 "Statistical Mechanics" Camb. Univ. Press
 Gerver, A. 1934 Dissertation, Amsterdam.
 Keosom, W. H. 1903 *Commun. Phys. Lab. Leiden*, No 88.
 Michels, A. and Michels, C. 1935 *Proc. Roy. Soc. A*, **153**, 201
 Michels, A., Blassse, B. and Michels, C. 1937 *Proc. Roy. Soc. A*, **160**, 358.
 Michels, A., Michels, C. and Wouters, H. 1935 *Proc. Roy. Soc. A*, **153**, 214.
 Nijhoff, G. P., Michels, A. and Gerver, A. 1930 *Proc. Acad. Sci. Amst.* **33**, 72.
 Otto, J., Michels, A. and Wouters, H. 1934 *Phys. Z.* **35**, 97.

The Isotherms of CO_2 in the Neighbourhood of the Critical Point and Round the Coexistence Line*

By A. MICHELS, B. BLAINSE AND C. MICHELS

(Communicated by F. A. Freeth, F R S.—Received 18 May,

Revised 11 August 1936)

The isotherms of CO_2 between 0 and 150°C . and up to 3000 atm. have been previously published by two of the authors (Michels, A. and C. 1935). The method used for these measurements was not suitable, however, for determinations in the neighbourhood of the critical point and the coexistence line. A second method has therefore been developed by which both the critical data and the coexistence line can be determined. This method and the results obtained are described in the present paper.

THE METHOD AND APPARATUS

The method was based on the one developed by Michels and Nederbragt (1934) for the determination of the condensation points of a binary mixture. While, however, for the measurements of condensation points, it was not necessary to know the quantity of gas in the apparatus, this knowledge is

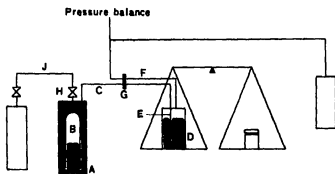


FIG. 1

essential for the determination of isotherms. A new apparatus was therefore constructed in which this quantity could be determined. A diagrammatic sketch showing the principle employed is given in fig. 1. In a steel vessel A, a glass bell B is suspended which is connected through the steel valve H and the capillary J to a cylinder containing a supply of the gas to be

* 50th publication of the van der Waals Fund.

examined. A steel capillary C connects A with a second steel vessel D , placed on one scale pan of a balance. Inside D a steel tube E , which is coupled to C , reaches to the bottom. The capillary F is connected to the top of D and leads to a cylinder of pure nitrogen and to an apparatus for measuring the gas pressure. The capillaries C and F are flexible, and are supported at G at such a distance from the scale pan that the variations in the forces acting on the latter during the swinging can be neglected. Before starting the measurements, the vessel A , the glass bell B and the tube C are completely, and the vessel D is partly filled with mercury. The valve H is then opened and CO_2 gas admitted to the glass bell, driving mercury out of A into D . The pressure in D is balanced by nitrogen introduced through F . When sufficient CO_2 has entered the glass bell, the valve H is shut. As the filling operation is carried out at a temperature and pressure at which the isotherms of CO_2 are known, the amount of gas in B can be calculated from a knowledge of the volume.

The pressure of the gas can be derived from the pressure of the nitrogen by applying a correction for the difference of the mercury levels in the glass bell B and the steel vessel D . The volume is equal to the volume of the mercury displaced and can be calculated from the increase of the weight of D . Here, however, three corrections have to be applied. one for the weight of the nitrogen in D , one for the expansion of the two cylinders by the pressure and one for the compression of the mercury.

As it is also necessary to know the temperature of the gas, the bomb A is placed in an oil thermostat. Having determined the quantity of gas in B , the isotherms can be measured by changing the pressure of the nitrogen and determining the new volume and pressure of the gas in B in the same way as has been already described. For measurements at other temperatures the bath in which bomb A is placed can be heated or cooled and the procedure repeated. In this case another correction must be applied for the thermal expansion of the steel and the mercury in A .

The thermostat was regulated at 25°C . and higher temperatures in the way described by two of the authors (Michels, A. and C. 1933, p. 416) and below 25°C . as described by Michels and Nederbragt (1934).

DETAILS OF THE APPARATUS

A more detailed drawing of the apparatus is given in fig. 2. The glass bell B , which had a volume of about 300 c.c., was sealed at K to a chromium steel tube, which was screwed, vacuum and pressure tight, into the top of the steel bomb A . The tube C , leading from the bomb in the thermostat to

the one on the balance, D , was divided into two parts by a small valve L . The lower part, C_1 , had an internal diameter of 2 mm., while the rest, C_2 , was a flexible capillary of 1.2 mm. outer and 0.8 mm. inner diameter. Preliminary experiments had shown that tubes of such diameters allowed pressure equilibrium to be reached within $1/1000$ of an atmosphere in a very short time. The length of C_2 from G , where it was supported, to the balance bomb was about 40 cm. The variation on the forces exerted by this tube on the scale pan of the balance during its swing was less than 50 mg.

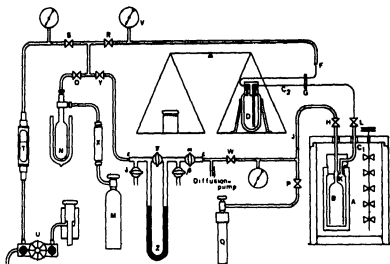


FIG. 2

The volume of the balance bomb was about 850 c.c. The nitrogen used for transmitting the pressure to the top of the mercury in it was supplied from the storage cylinder M . This nitrogen contained impurities amounting to 1 %, consisting of inert gases. As an extra precaution it was passed through a filter X and a stainless steel catchpot N , cooled with solid CO_2 . The nitrogen was admitted through the valves O and R and the capillary F to the balance bomb; this capillary F had the same dimensions as C_2 and was also supported at G . The pressure of the nitrogen could be measured to within 0.1 of an atmosphere by a spring manometer, V . For accurate measurements of the pressure, the valve S was opened, thus applying pressure on the oil surface in the glass levelling gauge T . This levelling gauge was coupled through the hydraulic press U to a pressure balance, with which the pressure could be determined to $1/1000$ of an atmosphere. The steel valves W and Y led to a mercury diffusion pump and a glass apparatus for the determination of some of the corrections.

DETERMINATION OF THE CORRECTIONS AND EXPERIMENTAL PROCEDURE

After the apparatus had been assembled, the balance bomb was filled with about 10.5 kg. of mercury. The valves O , S , L , and the glass taps β , and δ were shut and all the other valves and taps opened while the whole apparatus was evacuated with the diffusion pump. When a vacuum of about 1 cm. had been obtained, L was opened and the pumping continued for several days. Then the valves H and Y were shut and nitrogen admitted to the balance bomb by opening O , thus forcing mercury from this bomb into the thermostat bomb B . The pressure was then raised to 1 atm. L was shut and nitrogen again let into the balance bomb at about 30 atm. From the increase of weight, the temperature of the nitrogen and the known isotherms of this gas, the free volume of the balance bomb could be calculated. The procedure was repeated to 70 and 100 atm. Since the figures for the free volume agreed within the experimental accuracy (about 100 mg.) it was possible to neglect the expansion of the balance bomb with pressure. With a pressure in D of 100 atm. the valve L was opened. From the small decrease in the weight of D , the sum of the compression of the mercury and the expansion of the bomb A was calculated. This process was repeated at other pressures and it was found that the change in weight was proportional to the pressure within the experimental accuracy. From the results, the correction to be applied to the final measurements for the compression and expansion could be deduced. Pressure was released to 1 atm., the bath regulated at 50°C ., and from the increase of the weight of the balance bomb and the density of mercury at 25 and 50°C ., the thermal expansion of the bomb A was calculated.

The temperature was then again reduced to 25°C . and the difference between the mercury levels in A and D measured at different heights of the mercury in A , or at different weights of the balance bomb D , using the mercury differential manometer Z . This manometer was made of glass and was joined to the apparatus at E by soldered joints. The valve γ was shut and the balance bomb, together with the left-hand side of the manometer, opened to the atmosphere through δ (by opening R and Y). The other side of the differential manometer was connected with the capillary J (by opening α and W). A small amount of CO_2 was admitted from the supply bomb until the differential manometer showed a pressure of about 0.5 atm. The valve H was then opened and some CO_2 entered the glass bell till the pressure of the CO_2 was balanced by the difference in the mercury heights in B and D . This difference was then read directly from the levels of the mercury in Z . At the same time the weight of D was determined. By repeating this

procedure at different positions of Z , the difference between the levels in B and D as a function of the amount of mercury in the balance bomb was determined.

FILLING OF THE APPARATUS WITH CO_2

The valve L was shut and, as a precautionary measure, the CO_2 was pumped out of B with the diffusion pump till a high vacuum was obtained. W was then shut and L , P and Q opened. CO_2 , under the full pressure of the storage bomb, entered A , forcing the mercury into the balance bomb. When sufficient CO_2 had been let in, H was shut and the pressure counterbalanced in D by allowing nitrogen to enter the balance bomb. The nitrogen pressure was then measured with a pressure balance. It may be noted that a correction must be applied for the position of the oil level in the levelling gauge, which could be read to within 1 mm. The difference in the height of the mercury level in D and of the oil level in T caused a small hydrostatic head of compressed nitrogen which could be neglected. Applying the different corrections mentioned, the pressure of the CO_2 gas could be calculated from the load on the pressure balance, and the volume from the weight of the balance bomb.

The temperature of the thermostat was measured with mercury thermometers, divided in 0.01°C . and calibrated by the P.T.R. at Berlin. P , V and T being known, the amount of CO_2 and thus the normal volume, could be calculated from the isotherm data, published previously by two of the authors (Michels, A. and C. 1935). The determination of the normal volume was carried out at three pressures and the results agreed with each other to within $1/5000$. Two sets of measurements were carried out, the first with a normal volume of 9951 l. and the second with a normal volume of 20,612 l.

The volumes at other pressures and temperatures, including points in the liquid and coexistence region, were determined in exactly the same way.

The gas used was prepared from commercial CO_2 of "medical" quality* by a triple distillation in the way previously described by two of the authors (Michels, A. and C. 1933, p. 421). It was found that the vapour pressure during the condensation did not vary more than $1/20,000$, showing that the gas contained only minute quantities of impurities.

THE RESULTS

The results obtained are given in Table I, the isotherms measured with the normal value 20,612 l. being marked with an asterisk. The critical region

* The authors wish to thank the Ammoniakfabriek in Woesepe who kindly supplied the gas.

is given separately in fig. 3. Four isotherms, 0-027° C. below and 0-145, 0-280 and 0-483° C. above the critical temperature respectively, are plotted on a larger scale in fig. 4.

TABLE I

$T(^{\circ}\text{C})$	d	p	v	pv
2-85 ₂	52 288	36 318	0 0191248	0 69458
	53 159	36 663	0 0188115	0 68968
	53 852	36-932	0 0185694	0 68580
	54 751	36 992	0 0182645	—
	114 681	37 000	0 0087198	—
	433 67	37 000	0 0023059	—
	461-94	37-146	0-0021648	0 08041
	462 50	37 836	0 0021622	0 08181
	462 93	38 577	0-0021602	0 08333
	68 622	44-151	0 0150101	0 66271
10 82 ₂	68 019	44 810	0 0147018	0 65585
	69 483	45 053	0 0143920	0 64840
	74 705	45 256	0 0133860	—
	144 06	45 264	0-0069116	—
	370 30	45 262	0 0027005	—
	433 77	45 414	0 0023054	0 10470
	434 89	46 686	0 0022994	0-10735
	435 13	46 944	0 0022982	0-10788
	92 825	55 360	0 0107730	0 59640
	94 939	55 785	0 0105331	0 58759
19 87 ₄	97 346	56 224	0 0102726	0 57757
	99 042	56 261	0 0100967	—
	224 30	56 273	0 0044583	—
	342 56	56 270	0 0029192	—
	393 17	56 401	0 0025434	0 14345
	394-13	57 037	0 0025372	0 14472
	395 43	57 775	0 0025289	0 14611
	119-54	63-114	0 0083654	0 52797
	120 66	63 247	0 0082878	0 52417
	122 13	63 403	0 0081880	0 51915
25 07 ₀	124 42	63 451	0 0080373	—
	127 92	63-446	0 0078174	—
	137 91	63-450	0 0072511	—
	188 89	63-468	0-0052941	—
	290-88	63-467	0 0034378	—
	292-65	63 456	0 0034171	—
	359-82	63-474	0 0027792	0-17641
	360 00	63 513	0 0027778	0 17642
	360-31	63-593	0 0027754	0-17650
	360 99	63-760	0-0027702	0 17663
	362-45	64 151	0 0027590	0 17699
	366-24	65-259	0-0027305	0 17819
	366-72	65-370	0 0027269	0-17826
	372-83	67-514	0 0026822	0-18108

TABLE I (continued)

$T(^{\circ}\text{C})$	d	p	v	pv
25 07 ₀	118 08	62 946	0 0084688	0-53308
	120 58	63 231	0 0082933	0-52439
	121 95	63 377	0-0082001	0-51989
	126 04	63 450	0-0079340	—
	196-98	63-462	0-0050767	—
	329 42	63 459	0-0030356	—
	360 42	63 550	0 0027745	0 17632
	362 44	64 073	0 0027591	0-17678
25 29 ₈	119 42	63 284	0 0083738	0-52993
	123 95	63-750	0 0080678	0-51432
	*	124 34	0 0080425	0 51295
		124 65	0-0080225	—
	124-68	63-785	0 0080205	—
	125 11	63-793	0 0079930	—
	126 11	63 791	0 0079296	—
	324 45	63 798	0 0030821	—
	353 69	63 796	0 0028273	—
	357 82	63 812	0 0027947	0-17834
	358-53	63 975	0 0027892	0-17844
	359 13	64 124	0 0027845	0-17856
28 05 ₂	129 76	66 767	0-0077065	0-51454
	135-50	67 297	0 0073801	0 49666
	143 45	67 856	0 0069711	0 47303
	151 24	68 019	0 0066120	—
	194 85	68 015	0 0051322	—
	296-84	68-018	0 0033688	—
	332-66	68 128	0 0030061	0 20480
	336 72	68 597	0 0029698	0-20372
	341 73	69 350	0 0029263	0 20294
29-92 ₂	160 37	70 678	0 0062356	0 44072
	164-20	70 819	0 0060901	0 43129
	168 26	70 916	0 0059432	0 42147
	170 87	70 974	0 0058524	0 41537
	178 77	70 993	0-0055938	—
	205 87	71 000	0 0048574	—
	213 24	71 007	0 0046896	—
	232 93	71 003	0 0042931	—
	268 04	71-002	0 0037308	—
	296 13	71-002	0 0033769	—
	303 78	71-063	0 0032919	0 23393
	307-15	71-148	0 0032557	0 23164
	311 03	71 326	0-0032151	0 22932
	313 21	71-443	0 0031928	0-22810
30 40 ₂	168 62	71-513	0 0059305	0 42411
	174 58	71-653	0-0057280	0-41043
	179-37	71-737	0 0055751	0-39994
	184-28	71-775	0-0054274	—
	221-34	71-780	0 0045179	—
	275-97	71-787	0 0036236	—

TABLE I (continued)

$T(^{\circ}\text{C})$	d	p	v	pv
30-40 ₂	293.09	71.834	0.0034119	0.24509
	298.91	71.953	0.0033455	0.24072
	306.06	72.232	0.0032673	0.23600
31-01 ₂ *	85.415	59.237	0.0117076	0.69353
	185.12	72.632	0.0054019	0.39235
	192.69	72.710	0.0051897	0.37734
	200.02	72.757	0.0049995	0.36375
	206.22	72.778	0.0048492	0.35292
	209.21	72.787	0.0047799	0.34791
	214.24	72.794	0.0046677	0.33978
	227.82	72.796	0.0043894	—
	258.35	72.797	0.0038707	—
	267.51	72.821	0.0037382	0.27222
	272.69	72.846	0.0036672	0.26714
	286.91	73.002	0.0034854	0.25444
	300.94	73.419	0.0033229	0.24396
	85.418	59.239	0.0117071	0.69351
31 18 ₂ *	85.423	59.322	0.0117065	0.69445
	187.00	72.894	0.0053476	0.38981
	200.77	73.016	0.0049808	0.36368
	211.68	73.055	0.0047241	0.34512
	220.16	73.067	0.0045422	0.33188
	227.88	73.078	0.0043883	0.32069
	237.23	73.085	0.0042153	0.30808
	247.96	73.094	0.0040329	0.29478
	259.13	73.111	0.0038591	0.28214
	272.27	73.171	0.0036728	0.26874
	286.25	73.345	0.0034935	0.25623
	85.437	59.327	0.0117045	0.69439
31 32 ₂ *	85.422	59.387	0.0117066	0.69522
	187.06	73.072	0.0053459	0.39063
	200.66	73.207	0.0049836	0.36483
	211.65	73.260	0.0047248	0.34614
	220.30	73.281	0.0045393	0.33263
	227.80	73.294	0.0043881	0.32162
	237.24	73.307	0.0042151	0.30900
	247.88	73.319	0.0040342	0.29578
	259.45	73.347	0.0038543	0.28270
	272.51	73.426	0.0036696	0.26944
	286.31	73.622	0.0034927	0.25714
	85.426	59.389	0.0117060	0.69521
31-52 ₂	174.09	73.044	0.0057442	0.41958
	186.58	73.332	0.0053596	0.39303
	201.15	73.505	0.0049714	0.36542
	216.34	73.585	0.0046224	0.34014
	235.70	73.636	0.0042427	0.31241
	259.19	73.701	0.0038582	0.28435
	271.43	73.789	0.0036842	0.27185
	284.69	73.992	0.0035126	0.25990
	307.75	74.949	0.0032494	0.24354

TABLE I (continued)

$T(^{\circ}\text{C})$	d	p	v	pv
32.05 ₄	85 381	59 731	0 0117122	0.69958
	157 99	72 954	0.0063295	0.46177
	167.57	73 457	0 0059677	0 43837
	178 36	73 844	0 0056066	0 41402
	191 60	74 146	0.0052192	0.38699
	208.45	74 361	0 0047973	0.35674
	224.61	74.478	0 0044522	0.33159
	245 84	74 586	0.0040680	0.30339
	260 80	74 807	0 0037065	0.27727
	289 54	75 269	0 0034538	0 25996
	307 08	76 216	0 0032565	0.24820
	85 397	59 736	0 0117100	0 69951
34 72 ₁	85 341	61 008	0 0117177	0 71487
	115 14	69 709	0 0086851	0 60544
	123.59	71 408	0 0080913	0 57778
	132.87	72 969	0 0075262	0.54917
	132 91	72 969	0 0075239	0.54901
	144 07	74 459	0 0069411	0.51683
	156 02	75.690	0 0064094	0 48513
	171 91	76 858	0 0058170	0 44708
	189 49	77 726	0 0052773	0 41018
	210 80	78 420	0 0047438	0 37201
	237 00	79 066	0 0042194	0 33361
	259 80	79.691	0 0038492	0 30674
	287 37	80 982	0 0034798	0 28181
	318 82	83 484	0 0031366	0 26185
	85 352	61 011	0 0117162	0 71481
40 08 ₇	162 14	82 101	0 0061675	0 50636
	173 93	83.407	0 0057494	0 47954
	188 00	84 713	0 0053192	0.45060
	203 00	85 891	0 0049261	0.42311
	220 18	87.094	0 0045417	0 39556
	238 95	88 380	0 0041850	0.36987
	263 53	90 318	0 0037946	0 34272
	292 72	93 730	0 0034162	0 32020
	316 70	98 447	0 0031576	0 31085

To find the correlation between the isotherms obtained by the present method and those mentioned above, measurements were carried out at 40° C. and at densities within the density-range of the previous measurements.

The values of pv at the required densities were calculated from the previous measurements by interpolation. The results after correction for a small temperature difference are shown in columns 3 and 4 of Table II, while column 5 shows the differences. The mean difference is 1/5500, and the greatest is only 1 : 3000. This agreement can be considered good, as the method

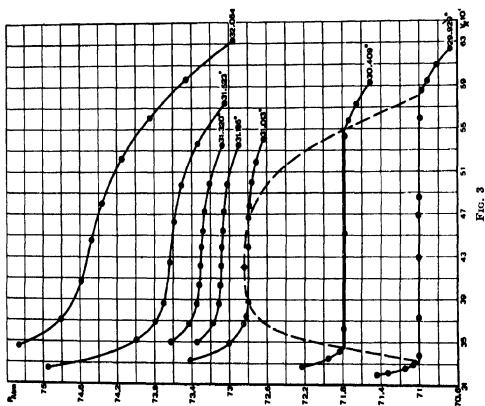


FIG. 3

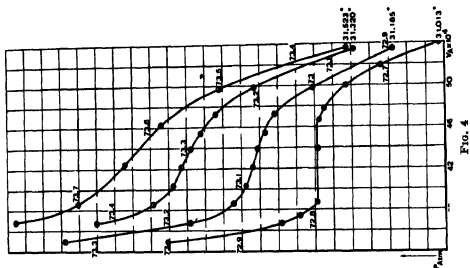


FIG. 4

described here, particularly for the higher densities, is not as accurate as the one used formerly. As pointed out above, however, the present method was developed for measurements in a region in which the previous method is inapplicable.

TABLE II

d	ρv (40° 087)	ρv (40°·105)	$\rho v_{\text{per ev.}}$	$A \times 10^6$
162 14	0·50636	0 50647	0·50638	9
173 93	0 47954	0 47966	0 47958	8
188 00	0 45060	0 45072	0·45066	6
203·00	0 42311	0·42324	0·42314	10
220 18	0 39556	0 39569	0 39575	-6
238·95	0 36987	0 37000	0 36708	-8
263 53	0 34272	0 34286	0 34293	-7
292·72	0 32020	0 32035	0 32043	-8
316·70	0 31085	0 31101	0 31091	10·5

DISCUSSION OF THE RESULTS

The Vapour Pressure

The vapour pressure of liquid CO_2 at the different temperatures can be obtained from the horizontal part of the isotherms. The results are given in Table III.

TABLE III—VAPOUR PRESSURE OF CO_2

t (° C.)	p (int. amt)	p_{calc}	$A = p_{\text{exp}} - p_{\text{calc}}$
2·853	36 997	36 997	0 000
10 822	45 261	45 409	-0·148
19·874	56 268	56 536	-0 268
25 070	63 456	63 736	-0 280
25·298	63·791	64 065	-0 274
28 062	68 017	68 146	-0·129
29·929	71 002	71·028	-0 026
30 409	71 780	71·778	+0 002
31·013	72 797	72 732	+0 065

The relationship
$$^{10}\log p = \frac{-875 \cdot 186}{t + 273 \cdot 15} + 4 \cdot 73909$$

can be used for interpolation. The deviation between the experimental values and those calculated from this formula are given in column 4 of Table III and are plotted in fig. 5. Meijer and van Dusen (1933) have published values for the vapour pressure of CO_2 ; to compare these values with ours, they have also been substituted in the formula and the deviations are plotted in fig. 5.

The Coexistence Line

The volumes of saturated vapour and liquid were determined from the intersection of the vapour and liquid curves with the horizontal part of the

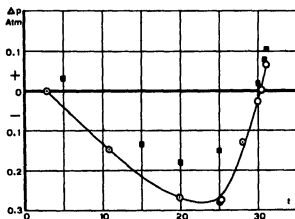


FIG. 5

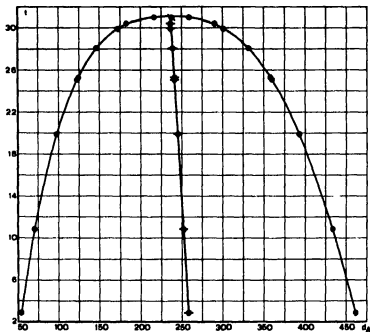


FIG. 6

isotherm. These volumes, and the corresponding densities, are given in Table IV. The densities are plotted as a function of temperatures in fig. 6.

TABLE IV—COEXISTENCE LINE OF CO₂

<i>t</i> (° C.)	<i>v</i> _{liq}	<i>v</i> _{gas}	<i>d</i> _{liq}	<i>d</i> _{gas}
2 853	0 002165	0 018506	461·8	54·04
10·822	0·002306	0 014244	433·6	70·21
19 874	0 002545	0 010244	393·0	87·62
25·070	0 002780	0 008152	359·8	122 67
25·298	0·002796	0·008034	357·7	124·47
28·052	0·003015	0 006828	331·7	146·46
29·929	0·003322	0 005804	301·0	172·30
30 409	0 003445	0·00549	290·3	182·22
31·013	0 00388	0·00462	259·0	216 4

The values of $\frac{d_{\text{vap}} + d_{\text{liq}}}{2}$ are also plotted in fig. 6. A straight line has been drawn in the diagram and it can be seen that, within the experimental accuracy, all the points lie on it showing that the empirical law of Matthias, for the rectilinear diameter, holds even in the neighbourhood of the critical point. The deviations of the experimental points from this line are given in Table V.

TABLE V—DEVIATIONS FROM RECTILINEAR DIAMETER

<i>t</i> (° C.)	$\frac{d_{\text{liq}} + d_{\text{gas}}}{2}$ (exp)	$\frac{d_{\text{liq}} + d_{\text{gas}}}{2}$ (calc)	<i>A</i>
2 85	257·93	257·93	0 0
10 82	251 92	251·97	-0·05
19 87	245·30	245 20	+0 10
25 07	241 21	241·31	-0 10
25 30	241 09	241 14	-0 05
28·05	239 07	239·07	0 00
29·93	236 66	237 67	-1 01
30·41	236 25	237 31	-1·06
31 01	237 7	236·9	+0·8

The values of $d_{\text{liq}} - d_{\text{vap}}$ could be represented by the well-known expression

$$d_{\text{liq}} - d_{\text{vap}} = C(T_k - T)^n,$$

except in the immediate neighbourhood of the critical point. Fig. 7 shows the difference between the experimental values and those calculated from the formula, with $C = 127·60$, $n = 0·357$. The formula, together with the deviation curve and the equation for the rectilinear diameter

$$\frac{d_{\text{liq}} + d_{\text{vap}}}{\alpha} = 260·07 - 0·7484t$$

give an easy method for the calculation of d_{liq} and d_{vap} at intermediate temperatures.

Lowry and Erickson (1927) have summarized all the data published on the liquid and gaseous densities of CO₂. In the temperature range above 23° C., the only published data are those of Amagat (1892) and two points on the gas side published by Cailletet and Matthias (1891). The agreement between these and the values calculated from the present figures is of the order of 3 %.

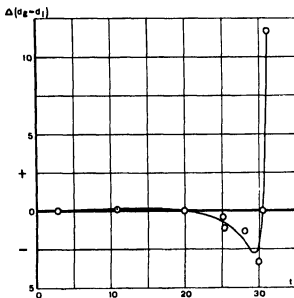


FIG. 7

The Latent Heat

The latent heat can be derived in the usual way from the slope of the vapour-pressure line and the volumes of saturated vapour and liquid using Clapeyron's equation

$$T \frac{dp}{dT} = \frac{\lambda}{v_2 - v_1}.$$

This latent heat can be divided into the internal latent heat and the work for the change in volume $p(v_2 - v_1)$. The results are given in Table VI and fig. 8.

It can be seen from the drawing that, in the immediate neighbourhood of the critical temperature, the values of λ fall rather abruptly to zero.

TABLE VI—HEAT OF VAPORIZATION

t ($^{\circ}$ C.)	(cal./mol.)	$P(v_{\text{gas}} - v_{\text{liq}})$ (cal./mol.)
2.85	2343.3	327.5
10.82	2047.5	292.5
19.87	1616.3	234.5
25.07	1269.3	185.6
25.30	1244.8	180.9
28.05	968.9	140.3
29.93	659.0	95.4
30.41	551.1	79.5
31.01	209.7	29.9

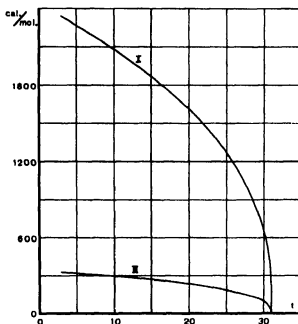


FIG. 8

Position of the Critical Point

The isotherms in the immediate neighbourhood of the critical temperature were measured with sufficient accuracy to allow the critical temperature to be determined from the inflexion points of the isotherms. The minimum values of $\left(\frac{dp}{dv}\right)_T$, found graphically from fig. 4, were plotted as a function of T ; within the experimental accuracy the graphical representation appeared to be a straight line. The intersection of this line with the T axis

gives the temperature where the tangent in the inflexion point is horizontal, i.e. the critical temperature.

Fig. 9 shows T_k to be $31.03^\circ \text{C.} \pm 0.01$.

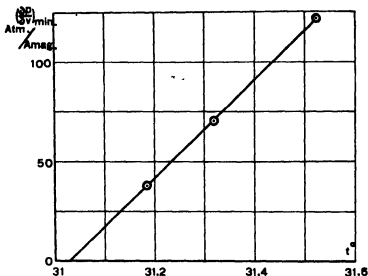


FIG. 9

In a similar way, by plotting the pressure of these inflexion points against p , the critical pressure is found to be 72.83^5 atm. Extrapolation of the vapour pressure line gives 72.82^5 atm. By extrapolation of the volumes of the inflexion points, the critical volume was found to be 0.00424 in Amagat units. Extrapolation however of the rectilinear diameter of Matthias gives the more reliable value of $d_k = 236.7$ Amagat units corresponding with a critical volume of 0.004224 Amagat units.

A check on the position of the critical point can be obtained from the formula mentioned above:

$$d_{liq} - d_{vap} = C(T_k - T)^n.$$

If this formula is extrapolated for $T = T_k$, both sides of the equation must become 0. The experimental values for $d_{liq} - d_{vap}$ lead to $T_k = 31.05^\circ \pm 0.01$.

The most probable value of T_k can be taken as 31.04°C.

The corresponding value of p_k then becomes 72.85 atm.

These figures are compared with those of other authors in Table VII.*

* The figures of the other authors have been taken from van Dusen.

TABLE VII—CRITICAL DATA OF CO₂

	t_c	p_c	V_c
Andrews, Th. (1869)	30.92	73	—
Amagat, E. H. (1892)	31.35	72.9	0.464
Kuenen, J. P. (1897)	31.1	73.26	—
Keesom, W. H. (1903)	30.98	72.93	0.440
Dorsman, C. (1908)	31.10	73.00	0.451
Cordoso, E. and Bell, T. M. (1912)	31.00	72.85	—
Meijer, C. H. and v. Dusen, M. S. (1933)	31.0	72.80	—
A. and C. Michels and Blaisse, B.	31.04	72.85	0.467

The Critical State

Several authors recently have expressed a doubt as to the ability of the continuity theory of van der Waals-Andrews to describe the phenomena in the neighbourhood of the critical point.

Jacyna (1935) developed a theory leading to an equation, called by him "the thermodynamical equation of state" from which he deduced that a critical region should exist rather than a critical point. On the boundary of this region the value $\left(\frac{\partial p}{\partial v}\right)_T$ should be zero. In the case of CO₂ this region should extend over a temperature range of about 0.43° C. The results as plotted in fig. 4 make it doubtful whether this theory is reliable.

Winkler and Maass (1933) published some papers on density discontinuities in the critical region; they concluded that density differences should persist above the critical temperature, which can be compared with a continuation of the two phases.

In some cases this leads to a picture of a coexistence line with a discontinuity above the critical point. The results given above do not support this suggestion, but it may be mentioned that similar phenomena are predicted by Kuenen (1907) and others if minute quantities of impurities are present.*

Ostwald (1933) attempted to give a qualitative description of the critical phenomena in the terminology of colloid-chemistry. From the results described above it may be concluded that for a qualitative description the concise theory of van der Waals suffices.

SUMMARY

In previous papers isotherm data of CO₂ have been published between 0 and 150° C. up to 3000 atm. The method used could not be applied in the

* Verschaffelt calculated that the presence of 0.1 % H₂ in CO₂ would lead to density differences of 24 % at a temperature of 0.5° C. above the critical point.

direct neighbourhood of the critical point and round the coexistence line. The present paper gives a description of a new method to measure isotherms in this region with an average accuracy of 1:5500, and the data obtained for CO_2 by this method.

The principle of the method is to compress gas with mercury and to determine the change of volume of the gas from the weight of the mercury displaced. This weight is measured without the mercury being taken out of the apparatus.

Besides the isotherm results, the vapour pressure curve, the critical data and the latent heat are given. The results do not support the doubts, which have been expressed lately in literature as to the real existence of a critical point.

REFERENCES

- Andrews, Th. 1869 *Philos. Trans.* 159, 11, 575
Amagat, E. H. 1892 *C.R. Acad. Sci., Paris*, 114, 1093.
Cailletet, L. and Matthias, E. 1891 *C.R. Acad. Sci., Paris*, 112, 1170.
Cordoso, E. and Bell, T. M. 1912 *J. Chim. Phys.* 10, 500
Dorsman, C. 1908 *Diss.* Amsterdam.
Jacyna, W. 1935 *Z. Phys.* 95, 253.
Keesom, W. H. 1903 *Comm. Phys. Lab. Leiden*, 88.
Kuenen, J. P. 1897 *Phil. Mag.* 44, 179
— — 1907 "Die Zustandsgleichung." Vieweg.
Lowry, H. H. and Erickson, W. R. 1927 *J. Amer. Chem. Soc.* 49, 2729.
Meijer, C. H. and v. Dusen, M. S. 1933 *Bur. Stand. J. Res.* 10, 381.
Michels, A. and Nichols, C. 1933 *Philos. Trans. A*, 231, 416.
— — 1935 *Proc. Roy. Soc. A*, 153, 201.
Michels, A. and Nederbragt, G. W. 1934 *Industr. Engng Chem. (Anal. ed.)*, 6, 135.
Ostwald, W. 1933 *Kolloidzchr.* 64, 50.
Winkler, C. A. and Maass, O. 1933 *Canad. J. Res.* 9, 613.
-

Thermodynamic Properties of CO₂ up to 3000 Atmospheres between 25 and 150° C.*

BY A. MICHELS, A. BLIJL AND MRS C. MICHELS

(Communicated by F. A. Freeth, F.R.S.—Received September 14, 1936)

In previous papers (Michels and Michels 1935; Michels, Michels and Wouters 1935) the results of isotherm measurements on CO₂ and a method for interpolation of the pv values at intermediate temperatures and densities have been published. From the data obtained, the specific heat at constant volume C_v , the free energy F , the total energy U , and the entropy S , have been calculated, and these results are given in the present communication.

The values of F , U and S at N.T.P. have been taken as zero. The values of C_v , F , S and U at a density of 1 Amagat unit ($\rho = 1$) have first been calculated for different temperatures. To the values, so obtained, has been added the increase of these quantities by compression. The values of C_v at $\rho = 1$ have been calculated, using the interpolation formula as given by Shilling and Partington (1928)

$$C_v = 6.688 + 0.00423t - 0.648 \times 10^{-6}t^2,$$

and the values of U , S and F from the equations

$$U_t = \int_0^t C_v dT, \quad (1)$$

$$S_t = \int_0^t \frac{C_v}{T} dT, \quad (2)$$

$$F_t = U - TS. \quad (3)$$

The effect of the compression has been calculated using the formulae

$$\left(\frac{\partial F}{\partial v}\right)_T = -p \quad \text{or} \quad \Delta F = \int_1^p \frac{p^2}{\rho^2} d\rho, \quad (4)$$

$$\left(\frac{\partial S}{\partial v}\right)_T = \left(\frac{\partial p}{\partial T}\right)_v \quad \text{or} \quad \Delta S = - \int_1^p \frac{1}{\rho^2} \left(\frac{\partial p}{\partial T}\right)_v d\rho, \quad (5)$$

$$U = F + TS \quad \text{or} \quad \Delta U = \Delta F + T\Delta S, \quad (6)$$

$$\text{and} \quad \frac{1}{T} \left(\frac{\partial C_v}{\partial v}\right)_T = \left(\frac{\partial^2 p}{\partial T^2}\right)_v \quad \text{or} \quad \Delta C_v = - \int_1^p \frac{T}{\rho^2} \left(\frac{\partial^2 p}{\partial T^2}\right)_v d\rho. \quad (7)$$

* 53rd publication of the van der Waals Fund.

ρ	25°	31.04°	40°	50°	75°	100°	125°	150°
0	— 7.5	— 11.5	— 18.9	— 31.2	— 64.3	— 111.9	— 171.5	— 242.4
40	+ 2052.9	—	—	+ 2226.8	+ 2385.6	+ 2530.9	+ 2663.5	+ 2774.2
80	2346.1	—	—	2566.9	2770.8	2960.1	3136.0	3299.5
120	2478.4	—	—	2731.5	2965.5	3184.2	3388.8	3580.7
160	—	+ 2620.3	+ 2720.5	2828.1	3086.8	3328.5	3555.8	3769.9
200	—	2668.9	2776.7	2893.4	3171.1	3432.3	3678.9	3912.1
240	—	2701.4	2815.6	2939.3	3244.4	3513.1	3777.0	4027.4
280	—	2724.7	2844.2	2974.2	3285.5	3580.2	3860.3	4127.0
320	—	2742.2	2866.8	3002.8	3329.4	3639.8	3935.6	4218.5
360	2667.3	—	—	3028.4	3370.3	3696.8	4007.8	4306.1
400	2677.8	—	—	3054.2	3412.0	3754.0	4081.4	4395.7
440	2690.7	—	—	3083.6	3458.1	3816.8	4160.9	4511.6
480	2709.8	—	—	3120.5	3512.8	3889.1	4250.8	4598.8
520	2739.7	—	—	3169.8	3581.0	3976.0	4356.0	4722.2
560	2786.2	—	—	3236.8	3668.4	4083.2	4482.6	4868.0
600	2855.0	—	—	3327.9	3781.1	4217.0	4637.0	5042.6

TABLE II—EXCESS OF ENTROPY OVER THAT AT N.T.P. IN CAL. T^{-1} /MOL.

ρ	25°	31.04°	40°	50°	75°	100°	125°	150°
1	+ 0.590	+ 0.727	+ 0.925	+ 1.148	+ 1.659	+ 2.148	+ 2.612	+ 3.054
40	- 7.270	—	—	- 6.641	- 6.077	- 5.550	- 5.059	- 4.570
80	- 9.198	—	—	- 8.474	- 7.857	- 7.181	- 6.780	- 6.299
120	- 10.559	—	—	- 9.712	- 9.046	- 8.400	- 7.921	- 7.425
160	—	- 11.354	- 11.004	- 10.678	- 9.978	- 9.357	- 8.820	- 8.309
200	—	- 12.219	- 11.828	- 11.487	- 10.766	- 10.176	- 9.588	- 9.069
240	—	- 12.928	- 12.530	- 12.187	- 11.492	- 10.838	- 10.279	- 9.759
280	—	- 13.520	- 13.153	- 12.821	- 12.109	- 11.486	- 10.927	- 10.409
320	—	- 14.072	- 13.744	- 13.426	- 12.730	- 12.112	- 11.564	- 11.041
360	- 14.913	—	—	- 14.033	- 13.351	- 12.739	- 12.407	- 11.672
400	- 15.509	—	—	- 14.661	- 13.988	- 13.377	- 12.825	- 12.312
440	- 16.160	—	—	- 15.326	- 14.656	- 14.043	- 13.490	- 13.020
480	- 16.872	—	—	- 16.039	- 15.364	- 14.746	- 14.183	- 13.660
520	- 17.648	—	—	- 16.805	- 16.118	- 15.489	- 14.915	- 14.383
560	- 18.486	—	—	- 17.626	- 16.921	- 16.275	- 15.686	- 15.146
600	- 19.389	—	—	- 18.503	- 17.774	- 17.107	- 16.530	- 15.948

TABLE III—EXCESS OF TOTAL ENERGY OVER THAT AT N.T.P. IN CAL./MOL.

ρ	25°	31.04°	40°	50°	75°	100°	125°	150°
1	+ 168.5	+ 209.6	+ 270.9	+ 339.7	+ 513.4	+ 689.7	+ 868.6	+ 1050.1
40	- 114.5	—	—	+ 80.7	+ 269.8	+ 459.9	+ 650.1	+ 840.4
80	- 396.3	—	—	- 171.6	+ 35.5	+ 237.2	+ 436.4	+ 634.2
120	- 669.9	—	—	- 406.9	- 183.7	+ 28.5	+ 235.0	+ 438.8
160	—	- 833.5	- 725.5	- 622.5	- 387.0	- 167.3	- 44.2	+ 253.9
200	—	- 1047.9	- 927.1	- 818.6	- 577.2	- 353.0	- 138.4	+ 74.4
240	—	- 1231.1	- 1108.1	- 999.0	- 756.5	- 531.4	- 315.6	- 102.3
280	—	- 1387.8	- 1274.6	- 1168.8	- 930.2	- 705.8	- 490.4	- 277.6
320	—	- 1538.3	- 1437.1	- 1335.8	- 1102.6	- 488.0	- 668.5	- 453.7
360	- 1779.0	—	—	- 1506.2	- 1277.7	- 1056.9	- 832.0	- 632.9
400	- 1946.2	—	—	- 1683.4	- 1458.0	- 1237.6	- 1025.0	- 814.2
440	- 2127.4	—	—	- 1868.9	- 1644.4	- 1423.5	- 1210.0	- 997.6
480	- 2320.7	—	—	- 2062.6	- 1836.1	- 1613.7	- 1396.5	- 1181.5
520	- 2522.0	—	—	- 2260.6	- 2030.5	- 1803.8	- 1582.2	- 1364.0
560	- 2725.3	—	—	- 2459.1	- 2222.6	- 1989.8	- 1762.6	- 1541.1
600	- 2925.7	—	—	- 2651.5	- 2406.8	- 2166.5	- 1944.6	- 1705.8

For interpolation at intermediate temperatures it may be worth mentioning that Table II gives the negative slope of the values given in Table I, as

$$S = - \left(\frac{\partial F}{\partial T} \right)_\rho.$$

Similarly, Table IV gives the slope of the values in Table III, as

$$C_\rho = \left(\frac{\partial U}{\partial T} \right)_\rho.$$

The values of U at temperatures 40 and 150° have been plotted as a function of ρ in fig. 1, and the values of ΔC_p , the increase of C_p on compression, in fig. 2.

TABLE IV— C_p IN CAL./MOL.

ρ/T	25°	31.04°	40°	50°	75°	100°	125°
1	6.79	6.82	6.86	6.90	7.00	7.11	7.21
40	8.30	—	—	7.75	7.60	7.60	7.60
80	9.65	—	—	8.75	8.20	8.05	7.91
120	11.85	—	—	9.70	8.70	8.40	8.20
160	—	14.3	11.60	10.45	9.10	8.65	8.40
200	—	16.8	12.45	10.90	9.25	8.85	8.50
240	—	17.3	12.60	11.00	9.30	8.85	8.55
280	—	15.4	12.00	10.75	9.25	8.85	8.55
320	—	13.0	11.20	10.35	9.15	8.75	8.50
360	12.05	—	—	9.95	9.05	8.70	8.50
400	11.50	—	—	9.70	9.00	8.70	8.50
440	11.20	—	—	9.65	9.00	8.75	8.50
480	11.15	—	—	9.65	9.10	8.80	8.61
520	11.20	—	—	9.75	9.25	9.00	8.75
560	11.45	—	—	10.00	9.45	9.20	8.95
600	11.80	—	—	10.30	9.75	9.50	9.20

DISCUSSION

The Internal Energy

If the van der Waals equation with constant a and b is correct, the internal energy should be a linear function of the density, as in this case $\left(\frac{\partial U}{\partial \rho}\right)_T = -a$.

If, however, the repulsive forces between two molecules act over a finite distance, then at high densities the energy will first decrease less rapidly, pass through a minimum for that density where the attractive forces are balanced by the repulsive forces, and finally increase at still higher densities.

The results as published show that the deviation from a straight line, although present, is not very pronounced, and only from the curves for $\left(\frac{\partial U}{\partial \rho}\right)_T$ (fig. 3) it can be seen that a minimum is to be expected at densities of about 800. Attention must be drawn to the fact that the general character of the energy-density curves for the different temperatures is the same, and that they do not show any singularity at the critical density.

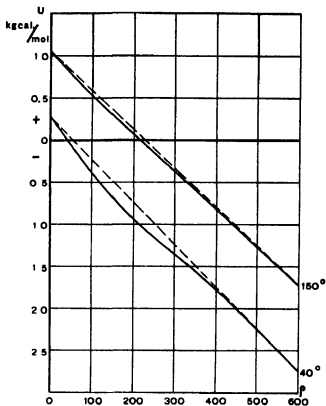


FIG. 1

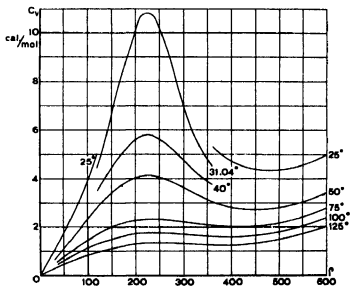


FIG. 2

The value of $\left(\frac{\partial U}{\partial \rho}\right)_T$ gives some information as to the internal pressure which can be defined by

$$p_i = T \left(\frac{\partial p}{\partial T} \right)_\rho - p$$

$$= -\rho^2 \left(\frac{dU}{d\rho} \right)_T.$$

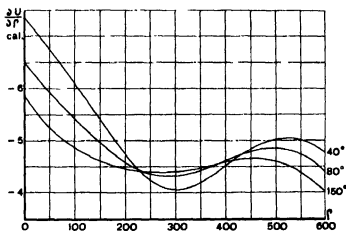


FIG 3

Some figures are given in Table V where p_i is the internal pressure, p_e is the external pressure and the "total" pressure p_t is given by

$$p_t = p_i + p_e.$$

TABLE V—INTERNAL AND EXTERNAL PRESSURE IN INTERNAL ATMOSPHERE

ρ	31.04°		50°	
	p_i	p_e	p_i	p_e
100	122	83.4	109	75.1
200	369	72.8	357	99.5
240	454	72.9	464	105.9
280	559	73.0	613	113.9
400	1305	118	1352	202
500	2321	407	2298	734
600	3335	1360	3224	1767

The Specific Heat

The character of the graphical representation of the specific heat and its change with temperature suggests that two independent phenomena are superposed. The first is a gradual increase of the specific heat with density, the second a specific influence of the critical density which fades out at the

higher temperatures. This suggestion is supported by the fact that the first phenomenon has a more general character and is also present in other gases far above the critical temperature (Michels, Wouters and de Boer 1936). The reason for this effect may be expected to be similar to that which causes the specific heat of a liquid always to be higher than that of the corresponding gas. This can be interpreted in the following way (Eucken and Seekamp 1928):

The potential energy of a molecule depends on its distance from other molecules, and therefore different configurations of molecules have a different energy and thus a different probability. At a given temperature the configurations with the lowest energy have the highest probability. With increase of temperature this probability decreases and more molecules go over to positions of higher potential energy. The energy required for this must be added to the ordinary specific heat, accounted for by the increase in kinetic energy. For dilute gases the configurations of low potential energy are few compared with those of high potential energy, and therefore the phenomenon is not pronounced. The higher the density, however, the fewer the latter, and the influence of the rearrangement of the molecules becomes more important. The fact that the additional specific heat at liquid densities (e.g. 600 A) is strongly dependent on the temperature (5 cal. at 25°, 2 cal. at 125°) shows that the phenomenon cannot be attributed to vibrations, as in solids. This, however, does not exclude the possibility that at still higher pressure the situation might become comparable with that of the solid state.

The remarkable maximum of C_p at the critical density, already found by Keesom (1908) for $T \sim T_K$, is of an entirely different character. The persistence of the position of the maximum at about the critical density through the whole temperature range, although clearly shown in fig. 2 at the lower temperatures, can be seen better for the higher temperatures from the character of the $\left(\frac{\partial p}{\partial T}\right)_\rho$ line as a function of T (fig. 2). The curve for the density 226, the experimental density closest to the critical one is, even at 150°, within the experimental accuracy, a horizontal straight line, the mathematical condition for a maximum of C_p .

The maximum of C_p at the critical density is not peculiar to CO_2 . Sydney Young (1895) has already found $\left(\frac{\partial^2 p}{\partial T^2}\right)_\rho$ to be zero for isopentane in the neighbourhood of the critical point. For a given equation of state the critical point is completely fixed by the conditions

$$\left(\frac{\partial p}{\partial v}\right)_T = 0, \quad \left(\frac{\partial^2 p}{\partial v^2}\right)_T = 0.$$

If $\left(\frac{\partial^2 p}{\partial T^2}\right)_T = 0$ holds for the critical density of any gas, it can therefore not be an independent condition, but must be a consequence of the equation of state.

The authors are not aware of any theoretical equation of state which fits this condition exclusively for the critical density. A kinetic explanation of the phenomenon might be based on assumption of group formation. The decreased association with increase of temperature would lead to a production of heat of dissociation and therefore to an apparent increase of C_v . A theory in agreement with the experimental facts should lead to the conclusion that $\left(\frac{\partial U_a}{\partial T}\right)_v$ is a maximum

at the critical density, where U_a is the energy of association; an assumption that the association is confined to the formation of double molecules, as was tried by G. v. Ry (1908), did not fulfil this condition.

The continuous transition of the critical isochore into the vapour-pressure line as described by Trautz and Ader (1934) cannot be regarded as an unexpected property of this isochore. The necessity of this phenomenon can be proved by thermodynamical reason as has been shown by Keesom (1908).

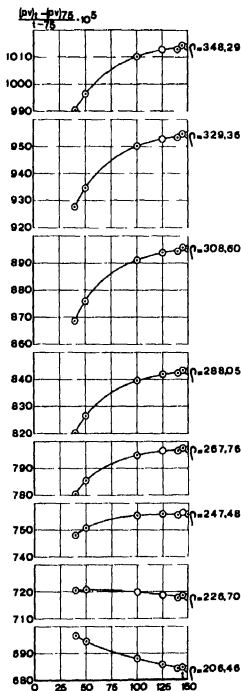


FIG. 4

REFERENCES

- Eucken, A. and Seekamp, H. 1928 *Z. phys. Chem.* **134**, 198.
Keesom, W. H. 1908 *Dissertation*, Leiden, p. 83. Compare also:
— Van der Waals-Kohnstamm, "Lehrbuch der Thermo-statik," **3**, 100.
Michels, A., Blaisse, B. and Michels, C. 1937 *Proc. Roy. Soc. A*, **160**, 358.
Michels, A. and Michels, C. 1935 *Proc. Roy. Soc. A*, **153**, 201.
Michels, A., Michels, C. and Wouters, H. 1935 *Proc. Roy. Soc. A*, **153**, 214.
Michels, A., Wouters, H. and de Boer, J. 1936 *Physica*, 's Grav. **3**, 597.
Ry, G. v. 1908 *Dissertation* Amst.
Shilling and Partington 1928 *Phil. Mag.* **7**, 1920.
Trautz, H. and Ader, H. 1934 *Phys. Z.* **35**, 446.
Young, S. 1895 *Proc. Phys. Soc. Lond.* **13**, 602.
-

The Mercury Photosensitized Exchange Reaction of Deuterium and Phosphine

By H. W. MELVILLE, *Laboratory of Colloid Science, Cambridge*, AND
J. L. BOLLAND, *Chemistry Department, University of Edinburgh*

(Communicated by James Kendall, F.R.S.—Received 7 November 1936)

In principle, the investigation of exchange reactions of deuterium atoms with hydrides is a simple problem. In practice, however, numerous difficulties arise in the experimental technique, and thus the method chosen for the study of any individual hydride must have regard to its chemical and physical properties. Now that a number of methods exist for investigating such reactions, it becomes of interest to attempt to find how such variables as the energy of activation and steric factor vary among the lower hydrides of the periodic system. The present paper deals with the kinetics of the exchange reactions involving phosphine and trideuterophosphine.

A static system was deemed most suitable for the present purpose, the atomic hydrogen being produced photochemically in the usual way.

APPARATUS

The essential features of the apparatus are indicated in fig. 1. The reaction vessel, *R*, a clear fused silica bulb of 110 c.c. capacity, was contained in an electric furnace, *F*, provided with a plano-convex silica lens, *S*, which served

to focus the light from a mercury lamp on the reaction vessel. The furnace temperature was measured by a calibrated platinum-platinum-rhodium thermocouple. A mercury manometer, a U-tube Apiezon oil B gauge, and a Pirani gauge were available for pressure measurement. The Pirani gauge was calibrated for each of the several gases used in the experiments. The mercury lamp was of the vertical type with tungsten anode and water-cooled mercury cathode; it was provided with a chromium-plated reflector. Exhaustion was effected by a three-stage mercury condensation pump.

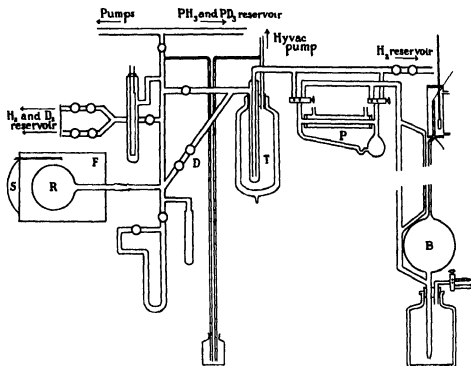


FIG. 1

Reservoirs containing hydrogen, deuterium, phosphine, trideuterophosphine and ammonia were sealed to the main vacuum line. The hydrogen and deuterium reservoirs were fitted with capillary pipettes to facilitate calibration of Pirani gauge and hydrogen analyser.

Preparation of Gases

Phosphine was prepared from phosphonium iodide and aqueous caustic soda. Trideuterophosphine was obtained by distilling deuterium oxide (98 %) vapour over calcium phosphide. Lump calcium phosphide, rapidly

pulverized and introduced into the apparatus, and then thoroughly baked out *in vacuo* (till evolution of gas ceased) was used for the purpose. Neglect of such precautions caused marked contamination of the resulting deuterophosphine with PH_3 . Traces of deuterium oxide and P_2D_4 were removed by fractional distillation. From a determination of the relative densities of 250 c.c. samples of PH_3 and deuterophosphine the deuterium content of the latter was estimated at $97 \pm 0.5\%$. Hydrogen from a cylinder was freed from oxygen by passage over heated palladized asbestos and through a liquid air trap. Deuterium was, for these experiments, prepared by completely decomposing trideuterophosphine on a well-outgassed tungsten filament. Synthetic ammonia from a cylinder was purified by repeated fractionation. Carbon dioxide was obtained by vaporizing solid CO_2 followed by fractionation.

Method of Analysis

The deuterium content of hydrogen mixtures obtained in the course of the experiments was determined from measurements of the thermal conductivity of the gas at high pressures (*ca.* 50 mm.) thereby eliminating the trouble often experienced in low pressure methods due to variable accommodation coefficients. In order to follow the course of the reaction throughout a run, it was most convenient to withdraw from the reaction

vessels samples of gas from time to time. It was accordingly necessary to reduce the size of these samples, and so that of the thermal conductivity cell, to the smallest practicable limits.* This was achieved by mounting a 0.015 mm. platinum wire in a capillary tube 50 mm. long and 2 mm. in diameter. With some care even narrower tubing might be employed. The technique for constructing such a gauge is shown clearly in fig. 2. To ensure that the wire be central in the capillary tube the 0.015 mm. wire is spot-welded to 0.3 mm. platinum wires of the shape shown at *a* and *b* in fig. 2. This assembly is then lowered into the capillary and the seal made at *c*. Maintaining the tube vertical the glass at *d* is sealed to *b* with a very small pointed flame. If necessary the wire may be tightened by softening the seal at *d* and

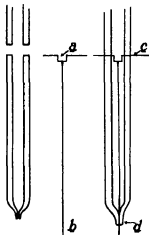


FIG. 2—Construction of hydrogen analyser

* The authors are indebted to Dr H. L. Roxburgh for developing the technique for constructing these small gauges.

pulling the wire *b* very gently downwards. To allow for the compression of the gas sample to be analysed in the thermal conductivity cell, the latter was affixed to a device resembling a McLeod gauge (cf. *B*, fig. 1) To maintain constant wall temperature the cell was surrounded with a water jacket which was supplied with water at about 20° C. from a thermostat. A thermometer reading to 0.05° C. gave the temperature of the cooling water.

Since normally the voltage across the gauge was about 10 it was essential to paint the external leads in order to prevent electrolysis occurring in the water jacket. It was found that introduction of a drop of Apiezon oil B on to the mercury in *B* prevented any tendency of the mercury to stick in the capillary. Samples of the reaction mixture, withdrawn by means of the small pipette *D*, were transferred to *B* by means of the small volume mercury condensation pump *P*. Since phosphine has a vapour pressure of about 0.007 mm. at liquid air temperatures it was necessary to free the mixture from its phosphine content by passage through a trap (*T*) maintained at about 70° K., at which temperature the vapour pressure of phosphine is only 10^{-5} mm. The effect of any residual phosphine on the resistance value of the hydrogen was then quite negligible. The pump *P* was connected to a system of taps (fig. 1) so that the sample after analysis might be returned to the reaction vessel.

The gauge wire was connected to a Wheatstone bridge, one arm of which was variable and read directly to 1 part in 10,000. A constant voltage of about 10 was applied across the bridge from a battery of large accumulators. Even at the pressures employed the thermal conductivity of hydrogen is not independent of pressure; and hence pressure adjustment of the gas confined in the conductivity cell had to be made accurately.

There were, then, three variables which might affect the resistance of the wire surrounded by a given sample of gas: (*a*) pressure of gas, (*b*) temperature of water jacket, and (*c*) voltage across the bridge. Arbitrary values of pressure and temperature were selected, and in practice resistance values were corrected for any slight deviations from these conditions by means of calibration curves.

One advantage of this method is that the sample of gas need not be of a predetermined volume, though to avoid isotopic separation it was essential to pump the whole sample from doser into the gauge. When this was completed the mercury was raised in *B* until the pressure in the closed limb was within 0.2 mm. of the arbitrary value—in these experiments 50 mm. The adjustment was facilitated by the use of a wire 50.0 mm. long. The difference in levels was then measured accurately (i.e. to 0.1 mm.) by a travelling microscope.

The gauge was calibrated with hydrogen-deuterium mixtures of known composition and also with similar mixtures in which the equilibrium $H_2 + D_2 \rightleftharpoons 2HD$ had been established photochemically by means of excited mercury atoms. The resulting calibration curves are shown in fig. 3. During such calibration experiments there was a possibility of slight voltage drift in the bridge battery. To eliminate small errors due to this the following procedure was adopted: duplicate resistance values of a sample of H_2 were taken, pressure adjustment being made before each. The resistance value of the hydrogen-deuterium mixture was determined similarly, and finally the resistance value of a second H_2 sample measured. The mean of the hydrogen resistance values was used to obtain the difference in resistance value (ΔR) between the mixture and pure hydrogen. Similar precautions against voltage drift were taken during all exchange experiments.

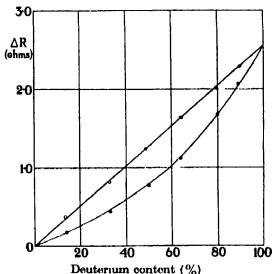


FIG. 3—Calibration curves of the hydrogen analyser for H_2-D_2 and equilibrium H_2-HD-D_2 mixtures ○ H_2-D_2 , ● H_2-HD-D_2 .

The estimated accuracy of the analytical method is 0.2 and 0.7 % for mixtures rich in deuterium and hydrogen, respectively. This could be improved upon by employing higher pressures in the gauge, though this would necessitate the use of larger samples of gas.

In order to show that the observed values of ΔR are reasonable, values of ΔR have been calculated from the simple theory, as follows. Let R be the resistance of the wire, at equilibrium temperature T , to which a voltage V is applied and which is surrounded by gas whose thermal conductivity

at temperature T , is $(\lambda)_T$. T_0 is the temperature of the walls of the containing capillary. In the stationary state,

$$\frac{V^2}{R} = k \cdot (\lambda)_T \cdot (T - T_0) + E,$$

where k is a constant and E represents the energy loss by conduction to the leads. From an application to platinum filaments of the method outlined by Langmuir, Maclane and Blodgett (1930), the difference in end-losses with H_2 and D_2 surrounding the filament, gave rise to an error of 0.05 ohm in the observed ΔR . E was accordingly neglected. If the surrounding gas were in turn H_2 and D_2 , at the same pressure and with the same voltage applied to the wire, then,

$$\begin{aligned} \frac{R_{D_2}}{R_{H_2}} &= \frac{T_{H_2} - T_0}{T_{D_2} - T_0} \cdot \frac{(\lambda_{H_2})_{T_{H_2}}}{(\lambda_{D_2})_{T_{D_2}}} \\ &= \frac{T_{H_2} - T_0}{T_{D_2} - T_0} \cdot \left(\frac{\lambda_{H_2}}{\lambda_{D_2}} \right)_{T_{H_2}} [1 - c(T_{D_2} - T_{H_2})], \end{aligned} \quad (1)$$

where c is the temperature coefficient of thermal conductivity of hydrogen. The temperature of the wire was given readily by the following relation, adapted from that given in *International Critical Tables*.

$$0.840 \log T = \log R - (\log R_0 - 0.840 \log T_0),$$

where R_0 is the resistance of the wire at temperature T_0 .

The value adopted for c was 2.50×10^{-4} , the mean of four determinations (Eucken 1913, Schleiermacher 1888; Schneider 1926; Gregory and Archer 1926).

Maass and van Cleave (1935) have obtained a value of 1.41 for $\left(\frac{\lambda_{H_2}}{\lambda_{D_2}} \right)_{T_{H_2}}$, in agreement with theory.

Taking $\left(\frac{\lambda_{H_2}}{\lambda_{D_2}} \right)_{T_{H_2}} = 1.41$, $R_{H_2} = 62.50$ and $T_{H_2} = 420.8^\circ \text{K.}$, and solving equation (1) graphically, $R_{D_2} = 66.0$ ohms, compared with the observed value of 65.3 ohms.

METHOD OF EVALUATION OF VELOCITY CONSTANT

The rate (R) of the exchange, say between deuterium atoms and phosphine, is expressed simply by the equation

$$-\frac{1}{2} \frac{d[D_2]}{dt} = R = k[D][PH_3].$$

The problem to be solved is the absolute evaluation of the bimolecular velocity constant, k . The apparatus described above permits the determination of the rate of production of hydrogen in the reaction mixture, i.e. $-\frac{1}{2} \frac{d[D_2]}{dt}$. $[PH_3]$ is known from the pressure and $[D]$ remains to be calculated. The stationary concentration of atomic deuterium is defined by the equation

$$\frac{d[D]}{dt} = \frac{I(2k_1[D_2] + k_2[PH_3])}{k_1[D_2] + k_2[PH_3]} - k_3[D]^2[X] = 0, \quad (2)$$

$$\therefore [D] = \sqrt{\frac{I(2k_1[D_2] + k_2[PH_3])}{k_3[X](k_1[D_2] + k_2[PH_3])}}, \quad (2a)$$

where I is the number of quanta absorbed in unit time by the mercury vapour and $k_1[D_2]$ and $k_2[PH_3]$ represent the rates at which D_2 and PH_3 deactivate excited mercury atoms, k_3 is the trimolecular velocity constant for the combination of atomic deuterium, and $[X]$ is the total pressure. It is assumed as a first approximation that PH_3 is as efficient as D_2 in promoting the combination of atomic deuterium. A further assumption is that all atoms combine in the gas. Later it will be shown that the majority of the atoms do combine in this manner. The term $k_2[PH_3]$ in the numerator of the first term of the right side of equation (2) is introduced, since a hydrogen atom is produced by the dissociation of the phosphine. This in turn immediately produces a D atom by the reaction $H + D_2 = HD + D$. Substituting equation (2a) in (1)

$$k = \frac{R}{[PH_3]} \cdot \left(\frac{I(2k_1[D_2] + k_2[PH_3])}{k_3[X](k_1[D_2] + k_2[PH_3])} \right)^{-1}.$$

The value of k_3 for hydrogen atoms with H_2 as the third body has been measured by Steiner (1935), who estimates it to be $13 \times 10^{15} \text{ cm.}^6 \text{ sec.}^{-1} \text{ mol.}^{-2}$. Amdur (1935) has measured k_3 for deuterium atoms with a D as third body and found it to be 1.4 times smaller than the corresponding hydrogen reaction. For D_2 as third body then k_3 will be taken to have the value $\frac{1.3}{1.4} \times 10^{16} \text{ cm.}^6 \text{ sec.}^{-1} \text{ mol.}^{-2}$.

The next quantity to be evaluated is I . The quantum input at 2537 Å during any particular run was determined by measuring the rate of the photosensitized decomposition of phosphine which had previously been calibrated against the uranyl oxalate actinometer (Forbes and Heidt 1934) in the following way. The reaction bulb was turned into a vertical position and a circular diaphragm of 1.5 cm. diameter carrying a 5 mm. thick silica

cell was interposed between lamp and bulb. With the filter cell filled with 50 % acetic acid solution to prevent direct photochemical dissociation of phosphine, the rate of mercury photosensitized decomposition of 25 mm. PH_3 was determined by a series of 30 sec. exposures. A fresh portion of the bulb was exposed for each period of illumination by rotation of the bulb from the ground-glass joint. The bulb was then cleaned and filled with actinometer solution (0.005 and 0.0025 M with respect to $\text{UO}_2\text{C}_2\text{O}_4$ and $\text{H}_2\text{C}_2\text{O}_4$ respectively) and exposed to the lamp for a time such that the decomposition did not exceed 10 %. Since the uranyl sensitized decomposition of oxalic acid is produced by light up to wave-lengths of 4500 Å, it was necessary to correct the above rate of decomposition for the contribution attributable to the quite considerable proportion of radiation of $\lambda > 2537$ Å. The experiment was accordingly repeated after substituting carbon tetrachloride for acetic acid in the filter cell, since CCl_4 absorbs 99.9 % of the 2537 Å lines (Melville and Walls 1933). The difference in the rates of decomposition was taken to be due to the 2537 line. (The extinction coefficient of the carbon tetrachloride decreases so quickly with increasing wave-length that strong lines of longer wave-length, e.g. 3030 Å, are not at all affected.) The correction to be applied for the reflexion of light at the silica-gas and at the silica-liquid interfaces was neglected as it amounted to only 2 %. The following are three typical results

TABLE I TEMPERATURE 20° C.

No of quanta entering R.V /sec	Rate of decomposition	γ_{PH_3}
4.3×10^{18}	1.20×10^{18} mol. sec ⁻¹	0.28
4.5×10^{18}	1.17×10^{18} "	0.26
1.41×10^{18}	0.37×10^{18} "	0.26

TABLE II—TEMPERATURE COEFFICIENT OF MERCURY-SENSITIZED DECOMPOSITION OF PH_3

Temp. ° C.	Rate in mm. of H_2 /min. measured at 20° C.	
17	0.070	
309	0.076	
319	0.074	
400	0.064	Allowance made for slight thermal decomposition
472	0.068	

Next the temperature coefficient of the rate of decomposition was measured so that decomposition rates measured at high temperatures could be converted directly to quantum inputs (Table II). As will be seen the tem-

perature coefficient is unity, as was also found with the direct photo-decomposition (Melville 1933, p. 548).

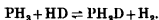
It remains now to evaluate the magnitudes of k_1 and k_2 . These constants contain the square of the quenching radii of D_2 and of PH_3 for excited Hg atoms. σ^2 for deuterium is known (Evans 1934, Zemanski 1930), and σ^2 for PH_3 and also for PD_3 was determined by a simple comparative method as 26.2 and 29.5×10^{-16} cm.² respectively (see following paper).

EQUILIBRIUM CONSTANT OF THE EXCHANGE REACTION

As will be seen later, it is not possible to follow the course of the exchange reaction to the equilibrium point on account of excessive photo-, and at high temperatures, thermal decomposition of the phosphine. It is therefore necessary to calculate the equilibrium concentration of deuterium in the hydrogen. Classically, of course, the H/D ratio in the hydrogen and in the phosphine are identical, but, owing to the existence of zero-point energy and other factors, the equilibrium distribution deviates slightly from classical values. It can be shown (Farkas 1936, p. 27) that the distribution ratio to a first approximation is given by the equation

$$\frac{(H/D)_{Hy}}{(H/D)_{Ph}} = \frac{2}{3} \cdot K,$$

where K is the equilibrium constant of the reaction,



At temperatures above 300° K. and not exceeding 1000° K. the equilibrium constant is given by the equation

$$-\ln K = \frac{\Delta E_0^\circ}{RT} - \frac{3}{2} \ln \frac{M_{PH_2D} \cdot M_{H_2}}{M_{PH_3} \cdot M_{HD}} - \ln \frac{\bar{I}_{PH_2D} \cdot \bar{I}_{H_2}}{\bar{I}_{PH_3} \cdot \bar{I}_{HD}} + \ln \frac{s_{PH_2D} \cdot s_{H_2}}{s_{PH_3} \cdot s_{HD}},$$

where ΔE_0° is the change in internal energy in the reaction at 0° K., \bar{I} refers to the mean moment of inertia, and s refers to the symmetry number. s has the values 1, 2, 3 and 1 for PH_2D , H_2 , PH_3 and HD , respectively. For the calculation of symmetry numbers see, for example, Wirtz (1936, p. 126). In order to compute ΔE_0° it is necessary to find the zero-point energy of PH_3 and PH_2D . Following on Robertson and Fox (1928), Fung and Barker (1934) have further analysed the infra-red absorption spectrum of phosphine and give the following frequencies for the four fundamental

vibrations. Raman spectra data (Yost and Anderson 1934 and Delfosse 1934), though not so accurate, are in accordance with these values.

		Fung and Barker cm. ⁻¹	Yost and Anderson cm. ⁻¹	Delfosse cm. ⁻¹
()	ν_3	990	929	—
(⊥)	ν_4	1121	1115	—
()	ν_1	2327	2308	2327
(⊥)	ν_2	992	—	—

The infra-red absorption spectrum of PD₃ has been observed by Sutherland (1936), who makes the following assignment for the fundamental vibration frequencies. These are in excellent agreement with the data obtained from Raman spectra of liquid PD₃ by de Hemptinne and Delfosse (1935). There is some discussion as to the correct frequency for ν_2 . According

	Sutherland cm. ⁻¹	de Hemptinne and Delfosse cm. ⁻¹
ν_3	730	740
ν_4	790	807
ν_1	1670	1664
ν_2	(730)	—

to Fung and Barker (1934), ν_2 is almost equal to ν_3 . For the purposes of calculation ν_2 will therefore be taken as 730 cm.⁻¹. Therefore the zero-point energy of PH₃ is 10740 cal. and that of PD₃ 7897 cal. From Howard's calculations (1935) on the vibrational frequencies of NH₃D the zero-point energy difference—1710 cal.—between NH₃ and NH₃D is very nearly one-third of that—5490 cal.—between NH₃ and ND₃. The zero-point energy difference between PH₃ and PH₃D on the same premises is taken to be 1015 cal. This value is supported by employing the method given by Rosenthal (1935) to calculate the frequencies of PH₃D from those of PH₃. This gives a zero-point energy difference of 1100 cal. Taking the zero-point energies of H₂ and of HD to be 6137 and 5323 cal. respectively, ΔE_0^0 amounts to 201 cal. In order to calculate the moments of inertia of PH₃ and PH₃D the following dimensions were taken for these molecules. H-H 1.9×10^{-8} cm. and P-H 1.5×10^{-8} cm. Using these data, Table III has been constructed to

TABLE III

T (° C.) ...	0	100	200	300	400	500	600	700
K	1.37	1.24	1.18	1.13	1.10	1.08	1.06	1.05
	41.3	42.6	43.4	44.0	44.4	44.8	45.0	45.2

show the variation with temperature of the equilibrium constant and also the equilibrium deuterium content of the hydrogen, for equimolecular $\text{PH}_3\text{-D}_2$ mixtures.

According to the calculations of Sutherland and Conn (1936), the values of the frequency, ν_2 , of the phosphine molecules assumed above are much too small. By using the valence-force-field method of computation these authors calculate ν_2 for PD_3 as 1690 cm.^{-1} . A similar calculation gives ν_2 for PH_3 as 2350 cm.^{-1} (Howard 1935). This assignment slightly alters the calculations made above: the total zero-point energy of PH_3 is 14,500 and that of PD_3 10,300 cal. respectively. This makes ΔE_0° for the reaction $\text{PH}_3 + \text{HD} = \text{PH}_3\text{D} + \text{H}_2$ equal to 586 cal. Unfortunately, since ν_2 has not been observed in the infra-red adsorption spectrum, an unequivocal decision cannot yet be made with the assignment of this frequency.

CORRECTIONS FOR THE NON-IDEAL REACTION

The above considerations apply to the ideal exchange reaction in which the phosphine is not appreciably decomposed. In actual practice, however, there is appreciable decomposition since k_2 is nearly equal to k_1 . At low temperatures besides decomposition there is the additional complicating factor of the formation of a film of red phosphorus, while at high temperatures thermal decomposition of the phosphine takes place. In each exchange run appropriate corrections have to be applied.

At this juncture, the procedure in carrying out a typical exchange run may be briefly described. 25 mm. of phosphine were admitted to the reaction bulb and illuminated for a short time, the pressure of the resulting hydrogen being measured on the Pirani gauge. This immediately gave (see p. 390) the number of 2537 Å quanta entering the bulb. The mixture obtained on addition of 25 mm. deuterium was then illuminated for suitable periods. Following each illumination, after establishment of uniformity of composition by diffusion, a gas sample was withdrawn for analysis. The increase of pressure due to decomposition was most conveniently measured by using the oil manometer differentially, taking readings at $\frac{1}{2}$ min. intervals during each period of illumination. At the conclusion of each run a second measurement of the quanta entering the bulb was made in order to check the constancy of output of the lamp.

The temperature range over which the required exchange reaction might be studied was somewhat limited: at temperatures much less than 450°C. , rapid formation of a film of red phosphorus, very opaque to 2537 Å radiation, on the reaction bulb surface rendered the progress of the reaction incon-

veniently slow. At temperatures much in excess of 600° C. the contribution of thermal decomposition (and probably exchange) to the apparent exchange was sufficiently large to mask that attributable to the exchange process being studied. The method of treatment of results adopted as most convenient involved the determination of the initial rate at which exchange proceeded. This has an additional advantage in that the exchange so determined may be attributed to a single reaction with reasonable certainty. The observed exchange is attributable to photochemical and thermal decomposition of phosphine in addition to true exchange processes. The calculation of initial observed exchange was greatly facilitated by use of the following simple empirical equation connecting the observed D content (D_{Hy}) of the hydrogen in the mixture and the pressure (Δp_{Hy}) of hydrogen formed by phosphine decomposition,

$$\log D_{Hy} = \log D_{D_0} - k \Delta p_{Hy}, \quad (3)$$

where k is a constant dependent upon temperature. The reasons for the validity of such an empirical relationship are briefly these. In an ideal exchange reaction there is an approximately linear relationship between the log of the D content of the hydrogen and the time of reaction. Where there is a very small amount of decomposition of the hydride the extent of decomposition will be directly proportional to the time and hence the $\log D - \Delta p_{Hy}$ relation will also be linear. In the non-ideal case under consideration the effects of two factors cancel to yield the same linear relation. These factors are (a) owing to the decomposition of phosphine the D content of the hydrogen is decreased abnormally quickly, and (b) on account of the diminution in the amount of PH_3 available for exchange the D content will decrease less quickly than in the ideal reaction.

The initial rate of observed exchange (R_1) was accordingly accurately determinable from

$$R_1 = k \cdot D_{D_0} \cdot \Delta p'_{Hy},$$

obtained by differentiation of (3) above. Hence, evaluating k from (3),

$$R_1 = \frac{2.3 \log_{10} \frac{4}{3}}{(\Delta p_{Hy})_{75}} \cdot D_{D_0} \cdot \Delta p'_{Hy}. \quad (4)$$

where $\Delta p'_{Hy}$ is the initial rate at which hydrogen is formed by decomposition, D_{D_0} the D content of the deuterium introduced into the reaction bulb, and $(\Delta p_{Hy})_{75}$ the value of Δp_{Hy} when $D_{Hy} = 75\%$, as determined from the $\log D_{Hy} - \Delta p_{Hy}$ graph.

The initial contribution of the decomposition to the observed exchange is

readily given by $\Delta p'_{\text{H}_2}/p_{\text{D}_2} \times 100$, where p_{D_2} is the pressure of deuterium originally introduced into the reaction vessel.

The following typical case will serve for illustration.

REACTION MIXTURE, 50 MM. PH_3 AND 50 MM. D_2 .

FURNACE TEMPERATURE = 490°C.

Time of illum. min.	D_{H_2} %	$\log_{10} \text{D}_{\text{H}_2}$	Δp_{H_2} mm. Hg	$\frac{2 - \log \text{D}_{\text{H}_2}}{\Delta p_{\text{H}_2}}$
0	100	2.000	—	—
4	95.3	1.979	1.09	0.192
11	90.8	1.958	2.71	0.155
20	85.4	1.931	4.55	0.152
28	81.7	1.912	6.18	0.134
37	75.0	1.875	7.88	0.159
47	70.0	1.845	9.58	0.162
59	65.2	1.814	11.48	0.162

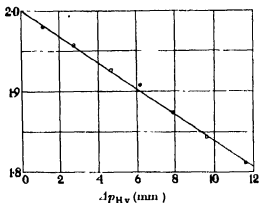


FIG. 4—A linear relationship was observed between the pressure of hydrogen, formed by decomposition of phosphine, and the logarithm of the deuterium content of the hydrogen in the exchange mixture.

From the satisfactorily straight-line $\log \text{D}_{\text{H}_2} - \Delta p_{\text{H}_2}$ graph (cf. fig. 4), when $\text{D}_{\text{H}_2} = 75\%$, $\Delta p_{\text{H}_2} = 7.85$ mm.

Hence from eqn (4), initial rate of observed exchange = 1.02% /min.

From a plot of $\frac{d\Delta p_{\text{H}_2}}{dt}$ and t , $\Delta p'_{\text{H}_2}$ was determined as 0.270 mm. H_2 /min.

\therefore Initial apparent exchange due to decomposition and back reaction effect (cf. p. 399) = $1.19 \times \frac{0.270}{50} \times 100 = 0.63\%$ /min

\therefore Initial rate of exchange = 0.39% /min.

The above-mentioned empirical equation seemed equally applicable whether either film formation or thermal decomposition were involved. In

the latter case, of course, allowance had to be made for the continuance of the thermal effect during the periods allowed for attainment of uniformity of composition after each illumination of the reaction mixture.

The intensity of quanta entering the bulb varied a little from run to run—owing mainly to non-reproducibility of location of the mercury lamp. It was accordingly necessary to investigate the relation between rate of exchange and input, since unless these variables were directly proportional some correction must necessarily be applied to the above-determined rates of exchange. With this object exchange runs were performed, in which the only parameter varied was the intensity. The intensity filter described by Melville and Walls (1933) proved suitable for the purpose.

Table IV gives the relevant data for three such runs in which the total pressure was 100 mm. and temperature 470° C., where x is defined by the

TABLE IV

Relative initial rate of exchange	Relative intensity	x
0.47	0.30	0.63
1.00	1.00	—
0.73	0.62	0.66

relation $R = kI^x$. These results would indicate a mean value of 0.64 for x . The deviation of the intensities used in the various runs from an arbitrarily chosen standard intensity, I_0 , was allowed for with the aid of the equation

$$R_0 = R \left(\frac{I_0}{I} \right)^{0.64}.$$

The above results give the necessary information for calculating the percentage of atomic hydrogen combining in the gas phase by a ternary collision mechanism. In the steady state the stationary concentration of deuterium atoms is defined by the equation,

$$\frac{d[D]}{dt} = 2I - k_4[D] - 2k_5[D]^2 = 0,$$

where the second and third terms account for removal of D atoms by diffusion to the walls and by combination in the gas phase respectively.

Since (cf. p. 389) R is proportional to $[D]$,

$$1/x = \frac{d \log I}{d \log [D]} = 1 + \frac{2k_5[D]^2}{k_4[D] + 2k_5[D]^2} = 1 + P,$$

whence

$$P = \frac{1-x}{x},$$

where P is the fraction of deuterium atoms recombining in the gas phase.

In these experiments then, this fraction was 0.57.

In Table V the results of a series of exchange runs in the temperature range 420–620° C. are detailed. The individual runs are illustrated in fig. 5. In each case, the exchange mixture consisted of 50 mm. PH_3 and 50 mm. D_2 .

TABLE V— PH_3 - D_2 EXCHANGE RUNS

Temp ° C	Initial rate of exchange	Initial intensity	R_0	γ_{EX}
420	0.17	0.156	0.13	0.21
422	0.13	0.147	0.11	0.19
470	0.40	0.121	0.34	0.50
490	0.39	0.120	0.32	0.47
560	0.79	0.102	0.78	1.11
620	1.30	0.104	1.25	1.64

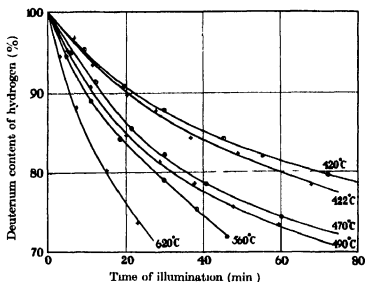


FIG. 5—Dependence of rate of exchange on temperature.

The "initial intensity"—expressed in mm. H_2 /min.—is the rate of production of H_2 in the phosphine decomposition run done as a preliminary to each exchange run. The standard intensity in the calculation of R_0 (cf. p. 397) was taken as corresponding to 0.100 mm. H_2 /min.

In a similar way, three typical runs for PD_3 - H_2 mixtures are detailed in Table VI.

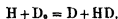
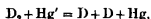
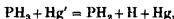
TABLE VI—PD₃-H₂ EXCHANGE RUNS

Temp. ° C.	Initial rate of exchange	Initial intensity	R ₀	γ _{Ex}
490	0.43	0.170	0.30	0.44
560	0.82	0.115	0.74	1.05
620	1.21	0.104	1.17	1.55

There is yet another factor to be considered besides exchange occurring by means of the chain-type mechanism



and so on, there is the possibility that some of the exchange, measured in the experiments described above, might be due to a back reaction, dependent on the primary dissociation of the phosphine



In order to find what contribution, if any, this process made to the total exchange rate, a mixture of 25 mm. PH₃ and 25 mm. D₂ was illuminated at 15° C., simultaneous measurements being made of pressure increase and the D content of the hydrogen gas. Excessive film formation on the reaction bulb militated against continuation of the run.

TABLE VII—PH₃-D₂ EXCHANGE RUN AT 15° C.

Observed D _{H₂}	94.8	89.8	86.1	83.5
Calculated D _{H₂}	95.6	91.6	88.0	85.6
	1.18	1.22	1.16	1.14
	1.19	1.23	1.18	1.16

Mean value = 1.19

In Table VII are tabulated the observed values of hydrogen D content and those calculated from the observed amount of decomposition. The third row gives the ratio of the observed to the calculated diminution in D_{H₂}. A small correction has been applied to this ratio (row 4) to allow for the fact that the D content of the atomic hydrogen is less than 100 % any recombination between H atoms and PH₂ radicals will of course not contribute to the observed exchange. It is assumed that owing to the rapidity of the reactions of the type H + D₂ = HD + D the D contents of the atomic and molecular hydrogen are the same. Since the temperature coefficient of

the photosensitized decomposition is unity it is extremely improbable that the back reaction effect would increase with increase of temperature. Accordingly, the contribution of the back reaction effect to the observed exchange was assessed at 0.19 time that due to decomposition (cf. pp. 395-6).

The quantum yields listed in Tables V and VI may therefore be attributed entirely to the chain mechanism given above. The fact that γ_{ex} exceeds unity at the higher temperatures is conclusive proof that this is the mechanism involved.

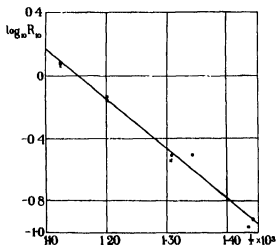


FIG. 6—The temperature coefficient of the exchange reactions, $D + PH_3 = H + PH_3D$ (⊙), and $H + PD_3 = D + PHD_3$ (×)

The slope of the satisfactorily straight-line $\log R_0 - 1/T$ graph gives an activation energy of 14.3 kcal. for the reaction $D + PH_3 = PH_3D + H$. The slope of the corresponding $H + PD_3$ line is practically identical. This value should, however, be corrected for the variation with temperature of the equilibrium value of D content of the hydrogen gas (cf. Table III). The correction involved is very small, however, amounting to only 0.1 kcal. The corrected value then for $D + PH_3$ is 14.4 kcal.

The vibrational frequency data, obtained from the valence-force-field computations (see p 394, above), would indicate a very similar correction: viz. 0.2 kcal.

Since it was essential to demonstrate that the quantum yield of the exchange reaction could exceed unity one run was made at 500 mm. total pressure at 510° C. The quantum yield rose to 1.2 from 0.6 at 100 mm. The reason for the small increase is due to the increased proportion of deuterium atoms combining in the gas phase as is evident from the following calcula-

tions. At low pressures and high pressures respectively $[D]$ is given by the equations (see p. 397)

$$[D] = \frac{2I[X]}{k_4}, \quad [D] = \sqrt{\frac{2I}{k_3[X]}}.$$

The value of $[D]$ at any pressure is the reciprocal of the sum of the separate probabilities of the D atoms disappearing by diffusion to the walls and by combination in the gas phase. Therefore,

$$[D] = \frac{1}{k_4/2I[X] + \sqrt{k_3[X]}/2I},$$

and hence,
$$\gamma_{\text{Ex}} = \frac{k[\text{PH}_3]}{k_4/2[X] + \sqrt{k_3[X]}/2I}.$$

Now at 100 mm. pressure 57 % of the atoms combine in the gas phase the interpolated value of γ_{Ex} being 0.6, and therefore from the observed value of the quantum yield at 500 mm. the percentage of D atoms combining in the gas phase is 96. Further runs were not made at high pressures since the duration of a run became unduly long and the method of calculation given on p. 397 eliminates the necessity of working at such high pressure to ensure complete gas phase termination of the chains.

COMPARISON OF PHOSPHINE AND AMMONIA RESULTS

Using the same apparatus, the rate of the exchange reaction of deuterium with ammonia has likewise been investigated. In this reaction, decomposition of the NH_3 is negligible and there is no disturbing thermal decomposition. The results are given in Table VIII.

TABLE VIII—EXCHANGE WITH AMMONIA. $\text{D}_2\text{-NH}_3$ 1:1 MIXTURE.
100 MM. TOTAL PRESSURE

Temp. ° C.	R_0 % D/min.	γ_{Ex}	γ_{Ex} Melville and Farkas
420	2.3	4.3	5.9
490	6.1	9.1	9.8
560	10.3	16.0	17.0

Activation energy 10.7 kcal.

It will be seen that the energy of activation 10.7 agrees very well with that (10) obtained, using a different experimental set up and method of analysis in which the para hydrogen conversion was used to measure the stationary hydrogen atom concentration (Farkas and Melville 1936). Even the

absolute rates of reaction are in excellent agreement as will be seen from the last two columns of the Table VIII, where the quantum yields of the reaction are given. That the reactions $D + PH_3$ and $D + NH_3$ possess similar steric factors is shown by the following comparison. The following figures show how the ratio of rates varies with temperature. Assuming that the steric factors of the two reactions are equal the difference in energy of activation is shown in the last column of Table IX. This amounts to 4.3 kcal. compared with $14.3 - 10.7 = 3.6$ kcal. calculated directly from the separate energies of activation. If anything the steric factor for the phosphine is smaller than that for the ammonia reaction.

TABLE IX

Temp ° C.	R_{D+PH_3} % D/min.	R_{D+NH_3}	Ratio	Additional energy of activation
420	0.13	2.3	17.7	4.0
490	0.32	6.1	19.0	4.5
560	0.78	10.3	13.2	4.3
Mean value				4.3

THE ABSOLUTE RATE OF THE EXCHANGE REACTION

A knowledge of the stationary concentration of atomic deuterium is necessary for the purpose of calculating the absolute rate of the exchange reaction $D + PH_3 = PH_2D + H$. Equation (2a) on p. 390, from which $[D]$ may be evaluated, on the assumption that the recombination of deuterium atoms occurred exclusively in the gas phase, must be somewhat modified to take into account the appreciable fraction of D atoms recombining on the walls.

$$\frac{d[D]}{dt} = \frac{I(2k_1[D_2] + k_2[PH_3])}{k_1[D_2] + k_2[PH_3]} - k_4[D] - 2k_3[D]^2[X] = 0.$$

where the term $k_4[D]$ takes into account the recombination at the walls. On p. 397 it was shown that $k_4[D] = 0.76(2k_3[D]^2[X])$. This equation holds only for the light intensity used in the present experiments. The variation in intensity, however, did not exceed $\pm 20\%$. Therefore, substituting to eliminate the unknown factor k_4 (unknown, since although the diffusion of D atoms to the walls could be calculated, the efficiency of the recombination at the walls is unknown),

$$[D] = \left(\frac{I(2k_1[D_2] + k_2[PH_3])}{3.52 \cdot k_3[X](k_1[D_2] + k_2[PH_3])} \right)^{1/2}.$$

The calculation of the bimolecular velocity constant, k , from the data of one particular experiment follows.

Composition of exchange mixture: 50 mm. PH_3 + 50 mm. D_2 .

Temperature = 560°C .

Quantum yield of exchange = 1.1.

Quantum input = 2.37×10^{-10} einstein $\text{cm.}^{-3} \text{sec.}^{-1}$.

Calculated $[\text{D}] = 1.04 \times 10^{-10}$ mol. cm.^{-3}

Initial rate of exchange = 2.52×10^{-10} g. atoms D $\text{cm.}^{-3} \text{sec.}^{-1}$

$$\text{Bimolecular constant, } k = \frac{R}{[\text{PH}_3][\text{D}]} \\ = 2.52 \times 10^6 \text{ mol.}^{-1} \text{ cm.}^3 \text{ sec.}^{-1}.$$

The number of collisions per sec per c.c. (N) between D atoms and PH_3 molecules was calculated from the usual formula:

$$N = N_1 N_2 \sigma_{12}^2 \left[8\pi RT \left(\frac{1}{M_1} + \frac{1}{M_{\text{PH}_3}} \right) \right]^{\frac{1}{2}},$$

where N_1 and N_2 are the number of molecules per c.c. and σ_{12} is the sum of the radii. σ_{PH_3} was calculated from the parachor and found to be 3.25×10^{-8} cm (cf. Sugden 1930, p. 32). The value of σ for atomic hydrogen, and therefore also deuterium, is 2.49×10^{-8} cm. (Hartek 1929). σ_{12} has been corrected for variation with temperature according to the Sutherland formula, $\sigma_T^2 = \sigma_0^2 \left(1 + \frac{\sqrt{C_1 C_2}}{T} \right)$, where C_1 and C_2 are the individual Sutherland constants for atomic deuterium and phosphine, namely 30.6 (Amdur 1936) and 271, calculated from the formula $C_s = 1.47 T_B$ (where T_B is the boiling-point of phosphine in $^\circ \text{K}$). σ_{12} at 560°K . is therefore 2.64×10^{-8} cm. The calculated value of N is accordingly 3.74×10^{18} collisions/c.c./sec. The number of atoms exchanging with phosphine initially is 1.53×10^{14} /sec. and therefore the collision efficiency is 4.40×10^{-5} . Now if the reaction $\text{D} + \text{PH}_3$ were of a simple bimolecular character having a steric factor of unity, the collision efficiency at 833°K ., corresponding to an energy of activation of 14.4 kcal., would be 1.74×10^{-4} . The ratio of these efficiencies is 0.25 and is usually termed the steric factor for this particular reaction.

In a similar way the steric factor for the ammonia exchange reaction, $\text{D} + \text{NH}_3 = \text{NH}_3\text{D} + \text{H}$, has been calculated taking the energy of activation as 10.7 kcal. The value obtained—0.33—agrees well with the corresponding phosphine steric factor, as is to be expected from the similarity of structure of the two molecules.

COMPARISON OF THE PH_3 AND PD_3 EXCHANGE REACTIONS

To obtain the difference in energies of activation of the reactions $\text{D} + \text{PH}_3$ and $\text{H} + \text{PD}_3$ two methods are available, (a) direct measurement of the activation energies, (b) comparison of the relative rates of the exchange reactions at the same temperature. As will be seen from fig. 6 (p. 400) the energies of activation are nearly the same within the experimental error, which is estimated to be ± 0.5 kcal. The following table shows the relative rates of exchange at three temperatures. These values must, however, be corrected for, owing (a) to non-classical distribution of the deuterium between the hydrogen gas and the phosphines (see p. 392), (b) to difference between stationary H and D atom concentrations, since D atoms combine 1.4 times less quickly (see p. 390), and (c) to difference in collision frequencies owing, mainly, to the difference in the masses of the H and the D atoms. With

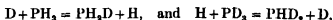
TABLE X

Temp. ($^{\circ}\text{K}$)	763	833	893
Observed ratio	1.07	1.05	1.07
Observed ratio, corrected for non-classical distribution	1.23	1.22	1.24
Observed ratio, corrected for stationary atom conc.	1.04	1.03	1.05
Observed ratio, corrected for mass factor	1.45	1.44	1.46
Difference in energy of activation*	570	610	660

* Calculated from the formula, Relative rate = $e^{-\Delta E/RT}$.

certainty, it may be said that the energy of activation of the reaction $\text{H} + \text{PD}_3$ does not exceed that of $\text{D} + \text{PH}_3$ by more than 600 cal. The difference in zero-point energy between the intermediate states will not be less than 2450 cal. This is what would be expected: for, to a first approximation, in the complexes, PH_3D and PD_3H , the energy associated with one H and one D atom will cancel, and therefore the zero-point energy difference should be about that between PH_3 and PD_3 which, according to the calculations, made on p. 393 is ca. 2000 cal.

From the above results, the absolute rates of the reactions,



are given by the respective equations,

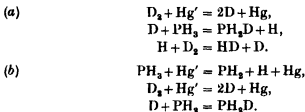
$$R_{\text{D-PH}_3} = 1.6 \times 10^{10} [\text{D}] [\text{PH}_3] e^{-\frac{14,400}{RT}},$$

and
$$R_{\text{H-PD}_3} = 2.5 \times 10^{10} [\text{H}] [\text{PD}_3] e^{-\frac{15,000}{RT}}.$$

SUMMARY

The mercury photosensitized exchange reactions of deuterium with phosphine and hydrogen with trideuterophosphine have been investigated from pressures of 100–500 mm. and at temperatures from 20 to 600° C.

Exchange occurs by two mechanisms:



Mechanism (b) is predominant at room temperature and has a temperature coefficient of unity, while (a) predominates at high temperatures owing to its having a high temperature coefficient.

By measuring the quantum input and calculating the velocity of combination of deuterium atoms the stationary atom concentration was calculated thereby permitting the evaluation of the bimolecular velocity constant of the reaction $\text{D} + \text{PH}_3 = \text{PH}_2\text{D} + \text{H}$. The energy of activation from the temperature coefficient is 14.4 kcal. and that for



The steric factor amounts to 0.2.

The equilibrium constant of the reaction $\text{PH}_3 + \text{HD} \rightleftharpoons \text{PH}_2\text{D} + \text{H}_2$ at different temperatures has been calculated from spectroscopic data.

The analytical method consisted in measuring the thermal conductivity of the hydrogen deuterium mixtures at high pressures, by means of a gauge having a volume of 0.2 c.c., the volume of gas required for analysis being 0.015 c.c. at N.T.P.

REFERENCES

- Amdur, I. 1935 *J. Amer. Chem. Soc.* **57**, 856–8.
 — 1936 *J. Chem. Phys.* **4**, 339–43.
 de Hemptinne, M. and Dolfosse, J. M. 1935 *Bull. Acad. Belg. Cl. Sci.* **21**, 793–9.
 Dolfosse, J. M. 1934 *Bull. Acad. Belg. Cl. Sci.* **20**, 1157–9.
 Eucken, A. 1913 *Phys. Z.* **14**, 324–32.
 Evans, M. G. 1934 *J. Chem. Phys.* **2**, 445–51.
 Farkas, A. 1936 *J. Chem. Soc.* pp. 26–35.
 Farkas, A. and Melville, H. W. 1936 *Proc. Roy. Soc. A*, **157**, 625–67.
 Forbes, G. S. and Heidt, L. J. 1934 *J. Amer. Chem. Soc.* **56**, 2363–8.
 Fung, L. W. and Barker, E. F. 1934 *Phys. Rev.* **45**, 238–41.

- Grogory, H. and Archer, C. T. 1926 *Proc. Roy. Soc. A*, **110**, 91-122.
 Harteck, P. 1929 *Z. phys. Chem. A*, **139**, 98-106.
 Howard, J. B. 1935 *J. Chem. Phys.* **3**, 207-11.
 Langmuir, I., Maclane, S. and Blodgett, (Miss) K. B. 1930 *Phys. Rev.* **35**, 478-503.
 Maass, O. and van Clove, A. B. 1935 *Canad. J. Res* **12**, 372-6.
 Melville, H. W. 1933 *Proc. Roy. Soc. A*, **139**, 541-57.
 Melville, H. W. and Walls, H. J. 1933 *Trans. Faraday Soc.* **29**, 1255-9.
 Robertson, Sir R. and Fox, J. J. 1928 *Proc. Roy. Soc. A*, **120**, 161-210.
 Rosenthal, (Miss) J. E. 1935 *Phys. Rev.* **47**, 235-7.
 Schleiermacher, A. 1888 *Ann. Phys., Lpz.*, **34**, 623.
 Schneider, E. 1926 *Ann. Phys., Lpz.*, **79**, 177-203.
 Steiner, W. 1935 *Trans. Faraday Soc.* **31**, 623-6.
 Sugden, S. 1930 "The Parachor and Valency," 1st ed. London: Routledge and Sons.
 Sutherland, G. B. B. M. 1936 Private Communication
 Sutherland, G. B. B. M. and Conn, G. K. T. 1936 *Nature, Lond.*, **138**, 641.
 Wirtz, K. 1936 *Z. phys. Chem. B*, **34**, 121-40.
 Yost, D. M. and Anderson, T. F. 1934 *J. Chem. Phys.* **2**, 624-7.
 Zemanski, M. W. 1930 *Phys. Rev.* **36**, 919-34.

The Photochemical Decomposition and Oxidation of Trideutero-phosphine

BY H. W. MELVILLE, *Laboratory of Colloid Science, Cambridge*,
 J. L. BOLLAND, *Chemistry Department, University of Edinburgh*,
 AND H. L. ROXBURGH, *Chemistry Department, University of Edinburgh*
 (Communicated by James Kendall, F.R.S.—Received 7 November 1936)

In the preceding paper, the kinetics of the exchange reaction of deuterium atoms and phosphine have been investigated. To enable the calculations to be completed in that paper it was necessary to determine the quenching radii of the PH_3 and PD_3 molecules for excited mercury atoms—and also the contribution to the total rate of exchange made by the back reaction, $\text{PH}_2 + \text{D} = \text{PH}_2\text{D}$. The first part of the present paper, therefore, deals with these two problems; the second part describes an analysis of the mechanism of the slow and explosive oxidation of trideutero-phosphine.

QUENCHING RADII OF PH_3 AND PD_3

The quenching radii of the phosphines have been determined by the following simple comparative method, making use of the absolute value of the quenching radius of hydrogen as determined by Zemanski (1930).

The arrangement of the quenching cell, F , spectrograph, S , and lamp, L , is shown in fig. 1. The standard gases selected were hydrogen and carbon dioxide. With the lamp running at a constant intensity, a series of exposures (20 sec.) was made, using different pressures of hydrogen in F , followed by an exposure with a determined pressure of phosphine admitted to F . The densities of the 2537 lines were measured on a photoelectric photometer and a plot made of the density against hydrogen pressure

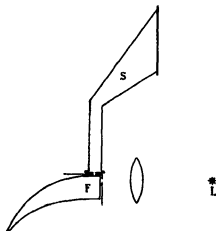


FIG. 1—Showing the arrangement of lamp, spectrograph and cell.

By inspection, the pressure of hydrogen giving the same density as that obtained when the tube contained phosphine, could be found. Under these conditions, if p_{H_2} and p_{PH_3} are the pressures of hydrogen and phosphine, σ_{H_2} and σ_{PH_3} the quenching radii, and M_{H_2} and M_{PH_3} the molecular masses, then

$$p_{\text{H}_2} \sigma_{\text{H}_2}^2 \left(\frac{1}{M_{\text{H}_2}} + \frac{1}{M_{\text{Hg}}} \right)^{\frac{1}{2}} = p_{\text{PH}_3} \sigma_{\text{PH}_3}^2 \left(\frac{1}{M_{\text{PH}_3}} + \frac{1}{M_{\text{Hg}}} \right)^{\frac{1}{2}},$$

and hence σ_{PH_3} may be readily calculated. In performing these experiments care was taken that the densities measured lay on the straight portion of the characteristic density-exposure curve of the photographic plate.

The following results (Table I) show the accuracy obtainable:

TABLE I

p_{PH_3} mm.	p_{H_2} mm.	Ratio	p_{PD_3} mm.	p_{H_2} mm.	Ratio
0.392	0.454	1.16	0.356	0.454	1.28
0.205	0.232	1.15	0.320	0.405	1.26
0.368	0.422	1.15	0.356	0.450	1.27
0.369	0.434	1.18	0.345	0.435	1.26

Mean 1.16

Mean 1.27

giving $\sigma_{\text{PH}_3}^2 = 26.2 \times 10^{-16} \text{ cm.}^2$ and $\sigma_{\text{PD}_3}^2 = 29.5 \times 10^{-16} \text{ cm.}^2$.

To test the applicability of this method, a determination of the quenching radius of CO_2 was made in a similar way. The value obtained, $3.8 \times 10^{-16} \text{ cm.}^3$, agreed well with the absolute value of $3.54 \times 10^{-16} \text{ cm.}^3$, due to Zemanski (1930).

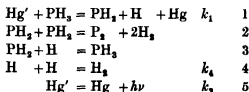
It will be noted that this method is especially suited for gases which are easily decomposed by excited mercury atoms, and which may produce an opaque deposit on the window of the quenching cell, since only one exposure is necessary to obtain a quenching radius value.

By using such a comparative method, troubles due to broad exciting lines and imprisonment of resonance radiation are obviated.

Excited mercury atoms ($^3\text{P}_1$) may be brought to the ground state, $^1\text{S}_0$, or to a metastable state, $^3\text{P}_0$, by collision with other atoms or molecules. A method of indicating which process occurs has been given by Zemanski (1930), who has shown that if there is a vibration frequency in the quenching molecules lying close to the energy of the transition, $^3\text{P}_1 \rightarrow ^3\text{P}_0$ (0.218 V), the quenching radius is invariably high. Thus the quenching radius of NH_3 is no less than three times that of ND_3 in virtue of the energy discrepancy of NH_3 (0.0160 V) being somewhat smaller than that of ND_3 (0.072 V). The energy discrepancies for PD_3 and PH_3 are approximately of the same order, namely 0.0119 and 0.069 V respectively, and hence it would be anticipated that PD_3 would quench much more efficiently (e.g. by a factor of three) than PH_3 , whereas in fact the efficiencies are very nearly the same. It would seem then that the quenching is to the ground state. Further evidence for this point of view will be given later.

THE PHOTSENSITIZED DECOMPOSITION OF PH_3 AND PD_3

Assuming the simple mechanism for the decomposition of phosphine previously suggested (Melville 1932), namely,



it is easily shown that

$$\frac{1}{R} = \frac{1}{R_\infty} \left(1 + \frac{k_3}{k_1[\text{PH}_3]} \right), \quad (1)$$

where R and R_∞ are the rates of decomposition at pressure, $[\text{PH}_3]$, and at infinite pressure, respectively. This of course assumes that the effect of the

secondary reactions on R is not influenced by variation in PH_3 pressure. Since the realization that imprisonment of resonance radiation plays an important part in the mechanism of such reactions (Melville 1935), the previous measurements (Melville 1932) have been repeated under such experimental conditions (i.e. low Hg pressures and narrow reaction tubes) as to eliminate this effect. The results for PH_3 and PD_3 are given in Table IV.

The values of k_1/k_3 calculated from the slope of the plot $\frac{1}{R} - \frac{1}{p_{1h}}$ given by these data are, for PH_3 0.280 mm.^{-1} , and for PD_3 0.425 mm.^{-1} . Taking the mean life of the excited mercury atom to be $1.07 \times 10^{-7} \text{ sec}$, the quenching radii have been calculated, from the usual collision formula, to be $5.8 \times 10^{-16} \text{ cm.}^3$ for PH_3 and $8.8 \times 10^{-16} \text{ cm.}^3$ for PD_3 , compared with the values of 26.2 and $29.5 \times 10^{-16} \text{ cm.}^3$ respectively, obtained by direct measurement. Two explanations of the discrepancy may be put forward: (a) the mechanism of the reaction involving the excited mercury atom is not quite so simple as that postulated above, and (b) the efficiency of the secondary reactions (2-4) varies with decreasing pressure in such a way as to reduce the quantum yield of the decomposition. The latter hypothesis can readily be tested by illuminating equimolecular mixtures of PH_3 and D_2 at various pressures, and taking simultaneous measurements of pressure increase and D content of the hydrogen. In general the amount of decomposition will be insufficient to account for the diminution in D content, in view of exchange proceeding by the back-reaction mechanism, indicated in the previous paper. A decrease in quantum yield, due to increased recombination of the primary decomposition products, will be reflected in an increase in the rate of "back-reaction exchange" relative to the rate of decomposition. The results presented in Table II indicate such a decrease in quantum yield.

TABLE II

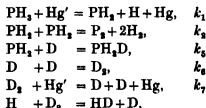
p_{PH_3} (mm.)	25.0	14.0	7.3	3.13	1.25	0.56
R_{EX}	5.2	4.0	3.6	2.7	1.59	1.17
γ	0.57	0.52	0.50	0.45	0.35	0.30

The data in the second row refer to the ratio (R_{EX}) of the rate of decomposition to the rate of exchange, attributable to recombination of D atoms and PH_3 radicals. These experiments were performed at 15°C. , when, of course, the exchange reaction $\text{D} + \text{PH}_3 = \text{PH}_2\text{D} + \text{H}$ takes place only to a quite negligible extent.

The quantum yield (γ) of the photosensitized decomposition of phosphine has been correlated with these observed ratios given in Table III in the

following manner. *As will be seen from the calculations, the method indeed proves to be a very accurate one for determining the quantum yield, the accuracy being much superior to that determined directly.*

The simple mechanism assumed to account for decomposition and exchange is



where k_5 and k_6 differ from k_3 and k_4 (on p. 408) by mass factors only, i.e. $k_5 = 1.4 \cdot k_3$ and $k_6 = 1.4 \cdot k_4$. The reasonable assumption, that the velocity of the reaction $\text{H} + \text{D}_2 = \text{HD} + \text{D}$ is sufficient to ensure removal of all H atoms formed in the dissociation of PH_3 , is made.

The stationary state equations involved are

$$\frac{d[\text{D}]}{dt} = 2K_1I - k_5[\text{PH}_2][\text{D}] - 2k_6[\text{D}]^2 = 0, \quad (2)$$

$$\text{and} \quad \frac{d[\text{PH}_2]}{dt} = K_2I - k_3[\text{PH}_2][\text{D}] - 2k_2[\text{PH}_2]^2 = 0, \quad (3)$$

$$\text{where} \quad 2K_1 = \frac{k_1[\text{PH}_3] + 2k_7[\text{D}_2]}{k_1[\text{PH}_3] + k_7[\text{D}_2]} \quad \text{and} \quad K_2 = \frac{k_1[\text{PH}_3]}{k_1[\text{PH}_3] + k_7[\text{D}_2]}.$$

$$\text{From equation (2),} \quad [\text{D}] = \sqrt{\frac{K_1(1-x)I}{k_6}}, \quad (4)$$

where x denotes the fraction of D atoms removed by recombination with PH_2 radicals.

$$\text{Let} \quad K_1(1-x) = K_3.$$

Substituting this value of $[\text{D}]$ in (3) and solving for $[\text{PH}_2]$

$$[\text{PH}_2] = \frac{\sqrt{I}}{4k_2\sqrt{k_6}} (\sqrt{(K_3k_3^2 + 8K_2k_2k_6)} - \sqrt{(K_3)k_6}). \quad (5)$$

The ratio of rate of decomposition of PH_3 to that of exchange, attributable to the "back-reaction" mechanism, is given by

$$R_{\text{ex}} = \frac{3(K_2I - k_3[\text{PH}_2][\text{D}])}{k_5[\text{PH}_2][\text{D}]}.$$

Inserting the values of $[PH_3]$ and $[D]$ obtained above, and making use of the relation

$$2\sqrt{(k_2 k_6)} = \sqrt{(1.4)} \left(\frac{\gamma}{1-\gamma} \right) k_6, \quad (6)$$

where γ is the quantum yield (i.e. the number of PH_3 molecules decomposed per excited Hg atom, deactivated by phosphine) of the photosensitized decomposition of phosphine (Melville 1935), the following relation between R_{Ex} and γ may be derived

$$R_{Ex} = 3 \left[\frac{1.4 K_2}{\sqrt{K_3}} \left(\frac{\gamma}{1-\gamma} \right)^2 \frac{1}{K_3 + 2.8 K_2 \left(\frac{\gamma}{1-\gamma} \right)^2 - \sqrt{K_3}} - 1 \right]. \quad (7)$$

For equimolecular PH_3 - D_2 mixtures, the quenching data on p. 407 give numerical values of 0.73 and 0.54 for K_1 and K_2 respectively.

It remains to evaluate K_3 . From equations (4), (5), and (6)

$$k_6 [PH_3] [D] = \left(\frac{1-\gamma}{\gamma} \right)^2 \frac{I}{1.4} \left[\sqrt{K_3^2 + 2.8 K_2 K_3 \left(\frac{\gamma}{1-\gamma} \right)^2} - K_3 \right]. \quad (8)$$

As a first approximation, x is taken as negligible: that is, $K_3 = K_1$. On this assumption, using a selected value of γ , $k_6 [PH_3] [D]$ may be calculated from (8) in terms of I . This value is inserted in (2), which is then solved for $[D]$. Comparison with (4) immediately gives a more accurate value of K_3 . Successive repetitions of this process, in each case using the value of K_3 obtained in the previous approximation, soon give a sufficiently stationary value of K_3 . A few corresponding values of R_{Ex} and γ are given in Table III.

TABLE III

γ	0.66	0.60	0.50	0.40	0.30	0.25
R_{Ex}	6.83	5.61	3.60	2.10	1.17	0.78

With the help of the resulting $\gamma - R_{Ex}$ plot the observed values of R presented in Table II have been used to give the value of γ (for the uninhabited decomposition) at the various phosphine pressures concerned. The dependence of γ on pressure of phosphine is shown graphically in fig. 2. The value of γ at infinite pressure (γ_∞), obtained by extrapolation of the $\gamma - 1/p$ curve, is 0.60.

This variation of γ with phosphine pressure necessitates slight modifica-

tion of the equation (1) on p. 408, giving the rate of mercury sensitized decomposition of phosphine, viz.

$$\frac{1}{R} \cdot \frac{\gamma}{\gamma_{\infty}} = \frac{1}{R_{\infty}} \left[1 + \frac{k_3}{k_1 [\text{PH}_3]} \right],$$

where γ is the quantum yield corresponding to phosphine pressure $[\text{PH}_3]$.

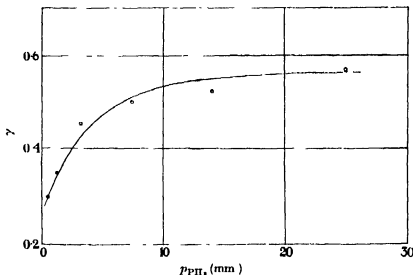


FIG. 2.—At low pressures, the quantum yield (γ) of the photosensitized decomposition of phosphine decreases, on account of change in the relative efficiencies of the various secondary reactions.

Table IV indicates the correction applied to the observed rates of decomposition (R) on account of the variable quantum yield factor.

TABLE IV

p_{PH_3} mm.	R mm./min.	γ	$\frac{1}{R} \cdot \frac{\gamma}{\gamma_{\infty}}$	p_{PD_3} mm.	R mm./min.	γ	$\frac{1}{R} \cdot \frac{\gamma}{\gamma_{\infty}}$
—	—	0.60	—	—	—	0.60	—
11.0	0.095	0.53	9.2	11.17	0.083	0.53	10.7
3.86	0.072	0.46	10.7	1.24	0.033	0.37	18.6
2.50	0.052	0.43	13.8	1.24	0.034	0.37	18.0
1.63	0.037	0.39	17.5	0.74	0.024	0.31	21.5
1.26	0.034	0.37	18.1				
1.17	0.030	0.36	20.0				
0.82	0.024	0.32	22.1				

In dealing with the PD_3 decomposition data, the same dependence of γ on phosphine pressure as obtained in the PH_3 case, was assumed.

A revised estimate of the quenching radii of PH_3 and PD_3 , based on the slope of the curve drawn between $\frac{1}{R} \cdot \frac{\gamma}{\gamma_\infty}$ and $\frac{1}{P_{\text{PH}_3}}$, gives $\sigma_{\text{PH}_3}^* = 17.3 \times 10^{-16}$ cm.² and $\sigma_{\text{PD}_3}^* = 24.9 \times 10^{-16}$ cm.². Obviously then, the variability of quantum yield can account satisfactorily for at least most of the discrepancy between the directly measured quenching radii and those calculated from variation of rate of decomposition with phosphine pressure (cf. p. 409), and this provides strong evidence that the primary reaction is simple dissociation of the phosphine molecule.

Since the values of σ for phosphine and hydrogen are known, it next becomes of importance to see if the inhibitory effect of hydrogen agrees with that calculated. Previous measurements (Melville 1932, p. 389) must be regarded as invalid, since the experimental conditions take no account of the imprisonment effect. In the experiments tabulated below, the pressures were chosen such that nearly all the mercury atoms excited were quenched by phosphine or hydrogen. This involved a serious experimental difficulty in measuring accurately the hydrogen formed by decomposition in presence of the somewhat high pressures of hydrogen required to produce appreciable inhibition. Further, the formation of the film of red phosphorus made it desirable to limit the extent of decomposition as much as possible. No great accuracy can accordingly be claimed for the value of the quenching radius derived from these results, namely 16.3×10^{-16} cm.². This is necessarily a minimum value, since the inhibition by competitive quenching is supplemented by that due to increased back reaction effect, $\text{PH}_3 + \text{H} = \text{PH}_2$.

TABLE V.—PRESSURE OF PHOSPHINE: 10 MM. TEMP. 18° C.

P_{PH_3} added mm	P_{PH_3} after exposure	Length of exposure min.	Mean P_{PH_3}	Rate mm/mm. R'	Uninhibited rate R_0	R_0/R'
—	0.155	1.0	0.077	0.155	0.155	—
0.632	0.775	1.0	0.70	0.143	0.154	1.08
0.309	0.451	1.0	0.38	0.142	0.153	1.08
0.663	0.861	1.5	0.76	0.132	0.151	1.14
1.020	1.272	2.0	1.15	0.126	0.149	1.18
—	0.148	1.0	0.074	0.148	0.148	—

Pressure of hydrogen required to reduce rate of decomposition to half uninhibited value = 5.4 mm. (extrapolated).

The agreement with the directly observed value of 26.2×10^{-16} cm.² is, however, sufficiently good to substantiate the supposition that phosphine quenches excited mercury atoms to the ground state; in the case of ammonia, where the excited mercury atoms are quenched to the metastable (3P_0)

state the discrepancy between the directly observed value of the quenching radius and that calculated from the inhibitory effect of hydrogen amounts to a factor of seven (Melville 1935, p. 331).

DIRECT PHOTODECOMPOSITION OF TRIDEUTEROPHOSPHINE

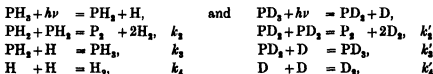
The direct photodecomposition of PD_3 has been compared with that of PH_3 . Differences were to be anticipated from two sources, namely differences in (a) quantum yield and (b) extinction coefficient. Since the effect of the latter factor would be apparent only at low pressures, when absorption is incomplete, any differences in decomposition rates observed at high pressures must be attributed to quantum yield differences. Using the same zinc spark intensity, the following comparison runs were made (cf. Table VI).

By plotting the reciprocal rate of decomposition against reciprocal pressure, the rates at high pressures (R_∞) were obtained by short extrapolations. It will be seen that PD_3 decomposes somewhat more rapidly than PH_3 (by a factor of 1.11). This value was supplemented by a series of comparisons of the rates of decomposition of PH_3 and PD_3 at pressures of 120 mm. On an average, PD_3 decomposed 1.10 times more quickly than PH_3 . [Individual results 1.05, 1.13, 1.10, 1.13.] This small difference in quantum

TABLE VI.—DIRECT PHOTODECOMPOSITION OF
 PH_3 AND PD_3 , AT 2100 Å (MEAN)

PH_3		PD_3	
Pressure mm.	Rate mm. PH_3 /min.	Pressure mm.	Rate mm. PD_3 /min.
∞	0.0880	∞	0.0966
116	0.0858	105	0.0940
18.9	0.0825	18.9	0.0740
9.2	0.0770	9.3	0.0562
4.6	0.0572	4.5	0.0358
2.08	0.0465	2.38	0.0234
0.96	0.0242	1.25	0.0176
0.40	0.0076		

yield, which must be attributed to a difference in the resultant efficiency of the secondary reactions, can be adequately accounted for in terms of the mass factors involved in the separate secondary reactions. Assuming the secondary reactions concerned are.



then it may be shown that

$$\frac{\gamma_{PD_3}}{\gamma_{PH_3}} \cdot \frac{1 - \gamma_{PH_3}}{1 - \gamma_{PD_3}} = \sqrt{\frac{k'_2 k'_4}{k_2 k_4}} \cdot \frac{k_3}{k'_3} \\ = 1.12,$$

on the assumption that the only differences in the velocity constants consist of differences in mass factor, i.e. $k_2 = 1.04 \cdot k'_2$, $k_3 = 1.36 \cdot k'_3$, $k_4 = 1.41 \cdot k'_4$.

Taking

$$\gamma_{PH_3} = 0.5,$$

$$\frac{\gamma_{PD_3}}{\gamma_{PH_3}} = 1.08,$$

in excellent agreement with the observed rates.

Similar comparisons, using the mercury lamp as source of radiation, gave a similar ratio of rates of decomposition, viz. 1.05. [Individual results: 1.13, 1.04, 1.02, 1.03, 1.07, 1.04.] At the pressures employed (120 mm.), it may safely be assumed that every mercury atom excited is deactivated by collision with a phosphine molecule, so that any difference in rate of decomposition must be attributed to differences in secondary reaction efficiencies and not in quenching radii. The equality of the above rate ratios for both direct and photosensitized reactions would seem to provide additional evidence that the secondary processes involved in both reactions are identical.

At low phosphine pressures, the higher extinction coefficient of PH_3 more than counterbalances the higher quantum yield of PD_3 , with the result that in this pressure range PH_3 decomposes more quickly than PD_3 .

The relative rates of direct decomposition and back-reaction exchange in 1:1 PH_3 - D_3 mixtures at 15° C. were determined in order to obtain estimates of the quantum yield at various pressures of phosphine. In these experiments it was found that the H_2 -HD- D_2 equilibrium was only partially established in the hydrogen samples withdrawn from the reaction vessel for analysis, following illumination: presumably the hydrogen formed by decomposition of the PH_2 radicals is of the molecular type, H_2 , and the atomic hydrogen concentration, which is much smaller than in the photosensitized case (by a factor of eight) is insufficient to produce equilibrium completely. Establishment of isotopic equilibrium in the samples of hydrogen analysed was accordingly completed catalytically, on a hot nickel filament.

R_{Kx} , Table VII, (as on p. 409) refers to the ratio of decomposition to back-reaction exchange—both expressed in terms of atoms of hydrogen—and γ is the quantum yield of the direct photodecomposition of PH_3 in absence of D_3 .

TABLE VII—1:1 PH₃-D₂ MIXTURES. TEMPERATURE = 15° C.

<i>p</i> (mm.)	10.3	4.6	1.45	0.26	0.22
<i>R_{EX}</i>	3.93	3.59	3.1	2.0	1.28
<i>γ</i>	0.52	0.47	0.36	0.27	0.25

The theoretical correlation of *R_{EX}* and *γ*, on which the values of *γ* given in Table VII are based, is much simpler in this case.

It is assumed that all H atoms, produced by photodissociation of phosphine, are replaced by D atoms by means of the reaction $H + D_2 = HD + D$.

Using the same notation for the secondary reaction velocity coefficients as on p. 410 the stationary state equations are

$$\frac{d[D]}{dt} = I - k_5[PH_2][D] - 2k_6[D]^2 = 0, \quad (1)$$

$$\text{and} \quad \frac{d[PH_2]}{dt} = I - k_5[PH_2][D] - 2k_2[PH_2]^2 = 0, \quad (2)$$

$$R_{EX} = \frac{3(I - k_5[PH_2][D])}{k_3[PH_2][D]}, \quad (3)$$

where *I* refers to the number of quanta, capable of dissociating a phosphine molecule, absorbed in unit time, as before.

$$\frac{2\sqrt{(k_2k_6)}}{k_6} = \sqrt{(1.4)} \frac{\gamma}{1-\gamma}. \quad (4)$$

It readily follows from equations (1)–(4) that

$$R_{EX} = 3\sqrt{(1.4)} \frac{\gamma}{1-\gamma},$$

$$\text{or} \quad \gamma = \frac{R_{EX}}{R_{EX} + 3.55}.$$

The value of *γ* at the higher pressures is in excellent agreement with that obtained earlier by a more direct method (Melville 1933). The identity of the quantum yields for direct and photosensitized reactions is convincing evidence that in both cases the primary dissociation products are a hydrogen atom and PH₂ radical.

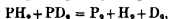
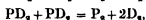
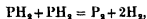
In connexion with the discrepancy between the photosensitized quantum yields as obtained using the uranyl oxalate actinometer (cf. previous paper) and from the above method, it may be pointed out that it has been found Farkas and (Melville 1936) that, compared with the chloroacetic acid actino-

meter, the uranyl oxalate actinometer may lead to quantum yields, too low by a factor of two. This matter is now under further investigation, as the discrepancy is rather serious.

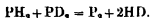
THE DECOMPOSITION OF PH_2 AND PD_2 RADICALS

While the non-establishment of the H_2 -HD- D_2 equilibrium, observed on illumination of D_2 - PH_2 mixtures with the zinc spark, may be taken as convincing evidence that the PH_2 radicals decompose in such a way as to give molecular hydrogen, it gives no indication as to the number of radicals involved in the reaction. Since the quantum yield of decomposition does not vary markedly with pressure, the secondary reactions are either unaffected, or affected in the same way by a change in total pressure of the phosphine. Moreover, since the number of atoms involved in the formation of a hydrogen molecule must be two, there is thus indirect evidence that the decomposition of the PH_2 radicals involves a collision of some kind between two such radicals. Inhibition by atomic hydrogen likewise supports this contention.

Confirmation of this view was obtained, when it was found that in the hydrogen, resulting from the direct decomposition of equimolecular PH_2 - PD_2 mixtures, the H_2 -HD- D_2 equilibrium was completely established. To explain this formation of HD by a mechanism independent of hydrogen atoms, it is necessary to assume that the decomposition of PH_2 radicals is a second order reaction, when, in addition to the molecular hydrogen formed by the reactions,



HD is produced by the reaction,



The above mechanism, which implies the formation of phosphorus in P_2 units, gives a simple explanation of the deposition of red phosphorus (Melville and Gray 1936).

THE KINETICS OF THE OXIDATION OF TRIDEUTEROPHOSPHINE

In previous papers (Melville and Roxburgh 1934) on the mechanism of the oxidation of phosphine, methods have been developed for measuring the

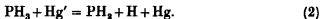
velocity coefficients of the elementary reactions comprising the whole process. These are

- (1) the rate of initiation of the reaction (direct or photosensitized),
- (2) the rate of propagation of the chains,
- (3) the probability of the chains branching,
- (4) the rate of termination of the chains,
 - (a) at low pressures, by diffusion to the walls of the reaction vessel,
 - (b) at high pressures, by destruction in the gas phase.

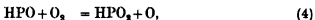
The determination of the rate of spontaneous starting of chains has not yet been accomplished but this is not important as it does not affect the explosion pressures of the $\text{PH}_3\text{-O}_2$ mixtures. Although it is possible to analyse the mechanism of the reaction in this way, only indirect information as to the identity of the chain carriers is available. In the case of the rate of initiation it is now certain that the primary step is



or

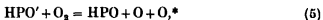


The speed of propagation is determined by the velocity of the following reactions



both reactions proceeding with practically identical velocities (Melville and Gray 1935).

Branching probably occurs thus



and termination on the walls involves



or



At high pressures, only that carrier which attacks the phosphine, namely the oxygen atom, is destroyed



X may be an oxygen, PH_3 , or inert gas molecule.

* The occurrence of this reaction is dependent upon the excited HPO molecule possessing sufficient energy to effect dissociation of O_2 , which requires 117 kcal. The heat evolved in reaction (3) is not known, but from the fact that 365 kcal. are evolved when 1 mol. of P_4 molecules is oxidized to P_4O_{10} , about 10 kcal. per gram atom will be evolved, on the average, by the reaction of one O atom with the partially oxidized molecule P_4O_n . It is therefore probable that energy of the order of 100 kcal. is liberated in reaction (3).

When trideutero-phosphine is substituted for phosphine, reactions (6), (7) and (8) will not be affected, for even in reaction (7) the rate of disappearance of HPO is governed by the rate of diffusion to the walls. Any differences found in the reactivity of the two phosphines can therefore be unambiguously ascribed to reactions (1)–(5). In so far as (1) and (2) are concerned, the data given in the first part of this paper permit the evaluation of the rate of the primary reaction.

THE STABLE REACTION ABOVE THE UPPER EXPLOSION LIMIT

First of all, the stable photo-induced reaction above the upper limit may be considered, thus involving reactions (1) or (2), (3), (4) and (8). Table VIII gives the results of two sets of experiments, one in which the zinc spark was employed, and the other, in which initiation was by excited mercury atoms.

At high pressures with a relatively large proportion of phosphine the chain length of the PH_3 oxidation is about 50 % greater than that for PD_3 , whereas at lower pressures and a smaller proportion of phosphine the chain lengths are identical. When the proportion of phosphine is high the ternary collision $\text{O} + \text{O}_2 + \text{PH}_3$ becomes of importance and it would therefore appear from the above results that this reaction is more efficient than $\text{O} + \text{O}_2 + \text{PD}_3$. When the pressure of O_2 is high, the predominating terminating collision is $\text{O} + \text{O}_2 + \text{O}_2$. Under these conditions, the chain lengths are identical and hence the propagation collisions $\text{O} + \text{PH}_2$, $\text{HPO}_2 + \text{O}_2$ and $\text{O} + \text{PD}_2$ and $\text{DPO}_2 + \text{O}_2$ must go at the same rate. To confirm this conclusion and also to measure the energy of activation of these reactions the chain length has been measured at the low temperature of -149°C . For this experiment the reaction vessel was completely immersed in a bath of pentane cooled with liquid air. Difficulty was experienced in maintaining constant temperature and the procedure finally adopted consisted in observing the temperature drift by means of the oil manometer, used differentially, before and after each period of illumination, so that a suitable correction might be applied. As will be seen from Table VIII (c), both reactions still went at the same speed though the chain length had fallen to 2.7. From this figure an upper limit to the activation energy of the propagation reactions may be calculated. This amounts to 620 cal. and consequently the efficiency of the collisions at 10°C , if the collisions are of a simple bimolecular character, is 0.33.

Having shown that the propagation reactions go with equal velocities, the probability of the chain branching may be compared by measuring the position of the upper explosion limit. These observations are plotted on a

logarithmic scale in fig. 3 where it will be noted that the two explosion curves are coincident. This implies that the reaction $\text{HPO}' + \text{O}_2 = \text{HPO} + 2\text{O}$ and $\text{DPO}' + \text{O}_2 = \text{DPO} + 2\text{O}$ also proceed with identical efficiencies. In fig. 4, the results for the lower limit are plotted in a similar manner. Again the explosion curves are coincident, demonstrating that the termination

TABLE VIII—STABLE REACTION ABOVE THE UPPER LIMIT

	P_{Ph}	P_{O_2}	Rate in mm. $\text{PH}_3/\text{min.}$	Chain length* (for $p_{\text{Ph}} = 10.0$ mm. and $\text{O}_2 = 100$ mm.)	
(a) Zinc spark.					
PH_2	10.6	—	0.0151†	15.6	
	11.0	101	0.255‡		
PH_2	9.12	—	0.0102	17.6	
	8.87	98	0.056		
PD_2	8.85	97.5	0.102	10.5	
	9.04	—	0.0102		
(b) Initiation by excited Hg atoms.					
PH_2	9.12	103	0.067	6.2	
	9.00	—	0.0127		
PH_2	9.00	100	0.067	7.3	
	9.16	—	0.0100		
PD_2	9.00	—	0.0150	3.9	
	9.00	100	0.053		
(c) Lower pressures. Zn spark.					
$p_{\text{PD}_2} = 0.254$ mm., $p_{\text{PH}_2} = 0.254$ mm., $p_{\text{O}_2} = 21.0$ mm., $p_{\text{O}_2} = 21.5$ mm.			$p_{\text{PH}_2} = 0.253$ mm., $p_{\text{PD}_2} = 0.261$ mm., $p_{\text{O}_2} = 22.0$ mm., $p_{\text{O}_2} = 22.0$ mm.		
t sec.	Δp_{PD_2} mm.	Δp_{PH_2} mm.	t sec.	Δp_{PH_2} mm.	Δp_{PD_2} mm.
10	0.007	0.007	10	—	0.001
30	0.014	0.015	30	0.003	0.003
60	0.018	0.021	60	0.005	0.005
90	0.023	0.023	90	0.007	0.008
120	0.025	0.025	120	0.009	0.012
Chain length	11	11	Chain length	2.7	2.7

* In a previous paper by Melville and Roxburgh (1934), the chain lengths quoted are too long by a factor of four, as is evident from the relative rates of oxidation and decomposition of phosphine. The probability of branching is likewise increased by the same factor and therefore amounts to 4×10^{-2} .

† Calculated from the equation, $2\text{PH}_2 = \text{P}_2 + 3\text{H}_2$.

‡ Calculated from the equation, $\text{PH}_2 + 2\text{O}_2 = \text{H}_2\text{PO}_2$.

$$\gamma_{\text{PH}_2} = 0.5, \quad \gamma_{\text{PD}_2} = 0.6.$$

efficiencies, i.e. diffusion of the chain carriers to the walls, and their destruction there, are equal.

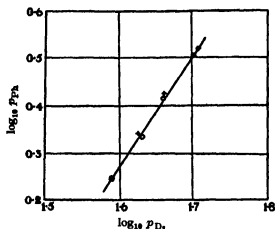


FIG. 3—The upper explosion limits for PH_3 and PD_3 are coincident.
+ PH_3 ; O PD_3 .

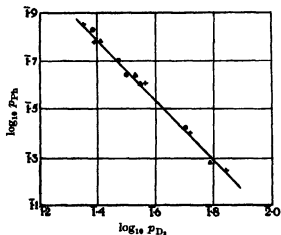


FIG. 4—The lower explosion limits for PH_3 and PD_3 are coincident.
+ PH_3 (ex PH_3I); O PD_3 ; Δ PH_3 (ex Ca_3P_2).

Finally, to check these conclusions, the chain lengths below the lower limit were measured in the manner described in earlier papers. The results quoted below would indicate that the chain lengths are the same within experimental error, thus substantiating the conclusions arrived at from the above experiments.

TABLE IX—CHAIN LENGTH BELOW LOWER LIMIT: ZN SPARK

PPH_2	PO_2	ν	PPD_2	PO_2	ν
0.429	0.430	85	0.430	0.430	91
0.428	0.430	77			

The above experiments provide the data for an approximate calculation of the reflexion coefficient of chains from a phosphorus oxide coated silica surface. According to the temperature coefficient of the chain length above the upper limit, the efficiency of the propagating collision is 0.33. Assuming that reaction occurs at every collision and that the reflexion coefficient from the walls is zero, the calculated value of the chain length is ten times that observed, implying that the collision efficiency of the propagating reaction is 0.1. Since these figures represent limiting values, the conclusion is that at temperatures of about 20° C. the chain carriers are destroyed immediately they collide with the wall.

The authors are indebted to Dr E. B. Ludlam for his very ready encouragement during these experiments, which were done in the Chemistry Department of the University of Edinburgh. They are also indebted to Professor E. K. Rideal for critical discussions during the progress of this work. One of them (H. W. M.) thanks the Commissioners for the Exhibition of 1861 for a Senior Studentship, during the tenure of which part of the work was carried out. The others thank the Trustees of the Moray Fund of Edinburgh University for grants toward purchase of apparatus.

SUMMARY

The direct and mercury photosensitized decomposition and oxidation of trideuterothiophosphine have been investigated with the following main results.

(a) The quenching radii of PH_3 and PD_3 for excited mercury atoms are 26.2 and 29.5×10^{-10} cm.² respectively. Evidence is adduced to show that the excited mercury atoms are quenched to the ground (1S_0) state.

(b) By means of determinations of exchange and decomposition in mixtures of D_2 and PH_3 at room temperatures, it has been shown that the extent of the back reaction $H + PH_3 = PH_2$ is consistent with a quantum yield at high pressures of about 0.5, for both direct and photosensitized decompositions of PH_3 .

(c) By similar means it has been demonstrated that the quantum yields for both reactions decrease at low pressures. This has been shown to account satisfactorily for discrepancies between the values of the quenching radii

of PH_3 and PD_3 obtained (a) by direct measurement and (b) from the variation of rate of decomposition with pressure of phosphine.

In the oxidation of PD_3 , it is shown that the following reactions proceed at similar velocities, since the position of both the lower and upper explosion limits coincide, and the chain lengths above and below these limits are identical.

Propagation of chain:



Branching of chain:



Termination at walls:



REFERENCES

- Farkas, A. and Melville, H. W. 1936 *Proc. Roy. Soc. A*, **157**, 625-51.
 Melville, H. W. 1932 *Proc. Roy. Soc. A*, **138**, 374-95.
 — 1933 *Proc. Roy. Soc. A*, **139**, 541-54.
 — 1935 *Proc. Roy. Soc. A*, **152**, 325-41.
 Melville, H. W. and Gray, S. C. 1935 *Trans. Faraday Soc.* **31**, 452-61.
 — — 1936 *Trans. Faraday Soc.* **32**, 271-85.
 Melville, H. W. and Roxburgh, H. L. 1934 *J. Chem. Phys.* **2**, 739-52.
 Zemanski, M. W. 1930 *Phys. Rev.* **36**, 919-34.

On the Values of Fundamental Atomic Constants

By STEN VON FRIESEN, D.Sc., *Uppsala*

(Communicated by O. W. Richardson, F.R.S.—Received 27 January 1937)

The practice of giving limits of error for results of physical measurements probably originates from a desire to indicate the narrowest region within which one is sure to find the true value of the quantity in question.

If it were possible to find a truly objective way of fixing this region, it would then be permissible to judge the value of an experiment from the breadth of the "region of error". It would also be possible directly to compare different determinations of a physical quantity.

The experimental errors, however, are of two principal kinds, accidental errors and systematic errors. Unfortunately, the former only can be treated in a quite objective way, that is, in this case, by mathematical methods. The systematic errors, on the other hand, can only be subject to an estimation where the experience and judgement of the experimenter himself must make the decision.

To avoid the difficulties of a decision of this kind many authors content themselves by giving some measure of their accidental errors and wholly neglect the systematic errors. They give, as a rule, the "probable error" as calculated by the method of least squares. In choosing such a course, they entirely abandon the thought of telling anything about the location of the true value. The probable error is a measure solely of the reproducibility of results obtained on a given occasion with one special instrument. It does not tell everything about the reliability of the method. The possibility of comparisons between results from different sources given by such a course of action is, therefore, purely imaginary.

To people who are not experimental physicists the inconveniences of such a practice are very severe. They lack those qualifications necessary for turning directly to the original publications to form an opinion of their own about the real accuracy. Such persons, as a rule, still look upon the limits of error as giving a region inside which the true value is to be found. As a result of this misconception many curious situations have arisen. We have had, at different times, among other things, one deflexion and one spectroscopic value of the specific electronic charge, one oil-drop and one X-ray value of the charge of the electron, and attempts have even been made to convince us that velocity of light is some sine function of time.

The course of action is dangerous even to the experimenter if the original papers do not give sufficient experimental details. In that case nothing but an intimate personal acquaintance with the author and his views can make a judgement of his work possible. One must, therefore, deeply regret those modern tendencies which suppress experimental details in order to reduce the bulk of scientific journals.

These questions were presented to the author of this paper in connexion with his experiments on c and h . It became necessary to form an opinion about the true accuracy of the values usually given for some fundamental atomic constants, and in some cases, to choose other values in better agreement with the results of modern developments in this field. Therefore, a survey was undertaken of the experimental work concerning the constants c , e/m , e , and h . The present paper gives an account of the results of this survey. The values given with their limits of error arise from attempts having been made to indicate the narrowest regions within which one is sure to find the true values.

THE VELOCITY OF LIGHT IN VACUUM

The velocity of light is, by far, the best known of our four constants. It has been the subject of a great number of determinations by eminent physicists. Table I contains the best of the determinations made during the

TABLE I—VELOCITY OF LIGHT

No.	Experimenter	Year	Method	c km./sec	Stated error	Light path	References
1	Cornu	1874	Fi	300,400	± 300	2×23 km.	(1876)
2	Michelson	1879	Fo	299,910	± 50	2×0.6 km.	(1879)
3	Newcomb	1882	Fo	299,860	± 30	2×3.7 km.	(1885)
4	Michelson	1882	Fo	299,853	± 60	2×0.6 km.	(1885)
5	Perrotin	1902	Fi	299,901	± 84	2×46 km.	(1902)
6	Rosa and Dorsey	1906	Indirect	299,781	± 30		(1906)
7	Mercier	1923	"	299,782	± 30		(1923)
8	Michelson	1924	Fo	299,802	± 30	2×35.4 km.	(1924)
9	Michelson	1926	Fo	299,796	± 4	2×35.4 km.	(1926)
10	Mittelstaedt	1928	Fi	299,778	± 20	$6-8 \times 41.4$ m.	(1929)
11	(Michelson)- Pease and Pearson	1932	Fo	299,774	± 11	$8-10 \times 1594$ m.	(1932)

last 60 years. Under the heading "Method" in the fourth column, information is given as to the general type of method used in the different cases.

Fi means that a system of periodic screening off of the light has been used. This system is best known in connexion with the name of Fizeau. Fo denotes Foucault's arrangement with rotating mirrors.

The most reliable of these eleven determinations might probably be the numbers 6, 9, 10, and 11.

Rosa and Dorsey (1906) obtained c by a comparison of the capacity of condensers in the units of the electrostatic and the electromagnetic system. The electrostatic capacity was calculated from the dimensions of the condensers, which were carefully measured. The capacity in electromagnetic units was obtained by measurements at the charging and discharging of the same condensers. c emerges as the square root of the relation between the capacities in the two systems. These determinations were made with the utmost skill and are a true model for precision measurements. The true accuracy, certainly, is one of the very highest ever attained at a determination of c .

In 1926 Michelson made very exact measurements with a rotating mirror between two mountains in California, Mt. Wilson and San Antonio Peak. The distance was 25,385.53 m. Unfortunately, it was not possible to measure it directly. A broken base line had to be laid in the immediate neighbourhood of the foot of the mountains. By the aid of this base the distance was found by triangulation. The probable error of the straight-line distance was 1 part in 6,800,000, that is, ± 5 mm. This corresponds to an error per metre of not more than 0.00015 mm. The officials from the U.S. Coast and Geodetic Survey, who made this determination, state that the actual error is surely less than 1 part in 300,000. It was their "feeling" that the actual error lay somewhere between 1 part in 500,000 and 1 part in 1,000,000. 1 part in 500,000 means $2\mu/m$. To make a measurement of length with this grade of precision in the ideal surroundings of a laboratory requires great care. Considering the fact that the field measurements have been made under very difficult conditions one is inclined to think this a rather bold statement. Other errors must be added to the error of distance, e.g. errors in the determination of the speed of the revolving mirror. It is obvious that the limits given (± 4 km./sec.) do not undoubtedly cover the true value region.

Mittelstaedt (1929) substituted Kerr cells for the toothed wheel used by Fizeau. This arrangement permits the use of a very high screening frequency, and is of great advantage as it allows for the use of very short distances which are easily measured with great accuracy. The limits of error seem to be quite reasonable.

Pease and Pearson's (1932) measurements took place in an evacuated

tube, 1 mile in length. The light was reflected several times at the ends of the tube. In this way the light was made to travel 8 or 10 miles. The experiment was a continuation of Michelson's work. To a certain extent the same experimental arrangements were used. The true accuracy is probably somewhat less than that shown by the limits of error.

These four determinations and Mercier's (1923) experiments on the velocity of electromagnetic waves on wires, from which c is calculated, give the velocity of light by four different methods. The agreement of the results is excellent. It might be said, with a high degree of probability, that the value of the velocity of light is now known with an error less than 1 part in 10,000. Nevertheless, some persons have thought it possible to conclude from the experimental results that the velocity of light is a sine function of time. Fig. 1, reproduced after de Bray, shows that the experimental

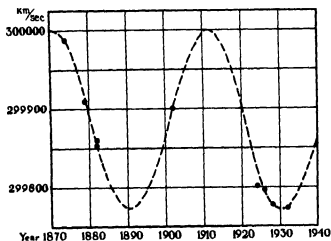


FIG. 1.—Variation with time of the velocity of light after E. J. Gheury de Bray (1934).

points really fit a sinus curve very well. If one turns, however, to fig. 2, where the limits of error are represented, it becomes quite obvious that the agreement is purely accidental. Moreover, the point of 1874 must be moved to a position not less than 410 km./sec. higher than shown by the diagram. de Bray used a value recalculated by Helmholtz which Cornu himself refused to accept. To state any variation in the velocity of light in virtue of the present material is entirely unjustifiable.

The result of our review is that

$$\text{the Velocity of Light in Vacuum} = 299,780 \pm 20 \text{ km./sec.}$$

THE SPECIFIC CHARGE OF THE ELECTRON (e/m)

A great number of physicists has undertaken measurements of e/m by various methods. The first determinations were made by Schuster and J. J. Thomson. They found values of the right order of magnitude, about 10^7 e.m.u./g. Since then the accuracy has steadily increased, although it is only lately that the reliability has reached a height to justify the use of the term "precision measurements". So recently as 1929 Birge thought it necessary to assume two different values of e/m , the deflexion value 1.769×10^7 e.m.u./g. and the spectroscopic value 1.761×10^7 e.m.u./g. The

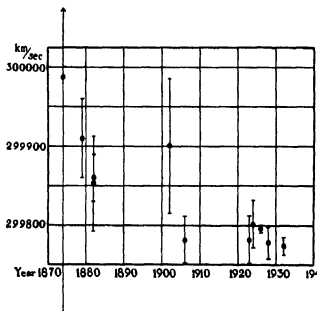


FIG. 2.—Determinations of the velocity of light, with stated limits of error.

first one was to be used in cases concerning free electrons and the other in cases concerning atomic structure. This difference does not exist in reality. It is due only to a considerable over-estimation of the accuracy attained at the time.

Table II gives a survey of the results of the most reliable measurements of e/m made during the last 10 years.

The determinations 1, 2, and 3 were made by a modification of a method first used by Wiechert. It consists in a direct determination of the velocity v of electrons, accelerated by a known voltage V , by measuring the time they require to travel a measured distance. At the beginning and the end of the

distance they pass through plate condensers to which the same alternating voltage of known frequency has been applied. The accelerating voltage is varied until the electron beam passes straight through the condensers without any deflexion. This means that the A.C. voltage is zero at both passages. Thus, a whole number of half-periods must have elapsed during the time required by the electron to travel the measured distance, and v is found easily. The formula $eV = \frac{mv^2}{2}$ gives e/m .

TABLE II—MODERN DETERMINATIONS OF THE SPECIFIC ELECTRONIC CHARGE

No.	Experimenter	Year	Method	e/m_0	Stated error	References
1	Perry and Chaffee	1930	Wiechert	1.761	± 0.001	(1930)
2	Kirchner	1931	"	1.7585	± 0.0012	(1932)
3	"	1931	"	1.7590	± 0.0015	(1932)
4	Dunnington	1933	Lawrence	1.7571	± 0.0015	(1933)
5	Houston (+ Camball and Kinsler)	1935	Zeeman effect	1.7570	± 0.0007	(1934, 1935)
6	Houston	1927	Rydberg constants H and He ⁺	1.7606	± 0.0010	(1927)
7	Shane and Spedding	1935	Rydberg ¹ H and ² H	1.7579	± 0.0003	(1935)

Dunnington (1933) used a method suggested by Lawrence. A filament emitted electrons which were accelerated by an alternating voltage of high frequency. The electrons describe circular paths under the influence of a magnetic field. A number of slits are arranged to define a circle of a certain fixed radius and those electrons which pass through the slits hit a collector. A cover round the collector has been electrically connected to the cathode. There is a value of the magnetic field at which no electrons are able to reach the collector. In this case, the time required by them to run through their circular path is equal to one whole period of the alternating voltage. They lose, therefore, their entire velocity and cannot enter the collector. Let θ be the path of the electrons in angular measure, ν the frequency and H_0 that value of the magnetic field for which a minimum number of electrons is registered at the collector, then $e/m = \frac{\theta \cdot \nu}{H_0}$.

The methods hitherto mentioned give the specific charge of free electrons. The next two methods concern electrons inside an atom. They are usually referred to as the spectroscopic methods.

The first of the methods depends upon the measurement of Zeeman splitting at a known strength of the magnetic field. Apart from minor corrections owing to the exact theory of the Zeeman effect, one finds $e/m = 4\pi c \frac{\Delta\nu}{H}$, where $\Delta\nu$ means the change of frequency effected by the application of the field H .

In the second spectroscopic method e/m is calculated from the Rydberg constants of light and heavy hydrogen or, of hydrogen and ionized helium. The formula used by Shane and Spedding (1935) for ^1H and ^2H will exemplify this method:

$$e/m = \frac{F(^2\text{H} - ^1\text{H}) \cdot R_{^1\text{H}}}{^2\text{H}(^1\text{H} - m)(R_{^1\text{H}} - R_{^2\text{H}})},$$

where F is Faraday's constant, ^1H and ^2H the masses of the atoms, and $R_{^1\text{H}}$ and $R_{^2\text{H}}$ the corresponding Rydberg constants.

In the first three determinations above there is a certain amount of uncertainty due to the possible occurrence of phase-displacements in the alternating voltages. Another cause of uncertainty lies in the unknown contact potentials which influence the accelerating voltage. Perry and Chaffee (1930) used concentrating coils to get a better intensity and definition of the electron beam. This arrangement must have an unfavourable effect on the accuracy. For this reason one seems justified in considering Kirchner's (1932) measurements somewhat more reliable than those of Perry and Chaffee, in spite of Kirchner's wider limits of error as given in his paper. With Dunnington's (1933) arrangement there seems to be little risk of phase-displacements and, furthermore, it is quite independent of contact potentials. Such an experiment, therefore, ought to be capable of giving very good values. All these four determinations (1-4) have one thing in common, viz. that they seem to have been made with the greatest skill and care. They give a mean value for the specific charge of free electrons $e/m = 1.759 \times 10^7$ e.m.u./g.

It is often said that spectroscopic determinations of e/m deserve particular credit as spectroscopic measurements can be made with a very high grade of accuracy. The actual position is such, however, that both with the Zeeman method and the Rydberg method the value of e/m depends upon small differences between spectroscopically measured magnitudes. Therefore, the accuracy obtainable is about the same with these methods as with the best methods working with free electrons.

Houston's value from the Zeeman effect is the outcome of a long series of careful determinations where special care has been devoted to the

measurement of the magnetic field. He gives limits of error which probably must be slightly increased. As regards Houston's other determination it seems advisable, even there, to widen the limits of error, among other reasons because of a certain degree of uncertainty as to which corrections must be used in consequence of fine structure in the spectral lines.

Shane and Spedding used the Rydberg constants of light and heavy hydrogen. These elements give lines with the same fine structure, which means that the errors are very much diminished. The errors given are probable errors. On account of a good agreement between results obtained with two different components of the lines the authors felt encouraged in the belief that the values are fairly free from systematic error. They felt justified in thinking that the probable errors describe closely the actual uncertainties in the results. However, the extreme values of e/m in their tables are 1.7566 and 1.7593. Therefore, it seems advisable to suppose that the real uncertainty is several times greater than the one given.

The above spectroscopic determinations give a value $e/m = 1.758 \times 10^7$ e.m.u./g., in very good agreement with the value for free electrons.

The determinations discussed above seem to permit the statement that the specific charge of the electron is now known with an accuracy of about one-tenth of a percent. They seem to indicate a value

$$e/m = (1.7585 \pm 0.002) \times 10^7 \text{ e.m.u./g.}$$

The figure 5 in the fourth decimal place is intended to tell that the best value seems to lie somewhere between 1.758 and 1.759, but it lacks any real physical significance.

THE ELECTRONIC CHARGE (e)

The first precision determination of the electronic charge was undertaken by Millikan in 1913. He used the famous oil-drop method. The measurements were preceded by extensive investigations into the method. The result obtained was verified by a new determination in 1917 which gave exactly the same result $e = 4.774 \times 10^{-10}$ e.s.u. The limits of error were given (1917) as $\pm 0.005 \times 10^{-10}$ e.s.u. Millikan's great renown and authority brought about the opinion that the question of the magnitude of e had practically got its definitive answer. A number of determinations by other people seemed to support this view.

In 1928 Bäcklin made precision measurements of the absolute wavelengths of some X-ray lines with a ruled grating. In this way he could find the factor of reduction between the conventional rock-salt scale of wave-

lengths and the optical scale. The knowledge of this ratio enables a calculation of e . The result was considerably higher than Millikan's and it was generally believed that some fundamental error was hidden in Bäcklin's method. His value was $e = (4.793 \pm 0.015) \times 10^{-10}$ e.s.u. Later on, the existence of a difference was confirmed by a number of new determinations by the same method. They were made by Bäcklin himself, Bearden in U.S.A. and others. The new experiments gave even higher values than Bäcklin's first determination. The method was scrutinized in many quarters in order to find the reason of the discrepancy.

It was suggested that the ordinary grating equation loses its validity at those very great angles of incidence which must be used in X-ray work. The suggestion never met with very great approval and lately it has been experimentally disproved. Another attempt to explain the discrepancy pointed out the possibility of a difference in density between large crystal blocks and the crystals actually taking part in the production of spectra. The large crystals should have a mosaic structure, that is, they should be built from a number of small crystals, more or less inserted into each other. Against this assumption strong arguments could be adduced, but nevertheless, Millikan's value of e was preferred generally. To arrive at a final decision it was necessary, either to repeat Millikan's measurements and find a fault there or, to develop a new independent method of an accuracy high enough to make the choice possible between the competing values of e . Both paths have been trodden during the last few years and both have given a decision in favour of the X-ray value, showing that the higher value must be the most nearly correct one.

In 1935 the author of this paper published results of a precision measurement of electron wave-lengths which enabled the calculation of e with an accuracy of about $\pm 0.1\%$. The wave-length was measured for electrons of known velocity. The voltage accelerating the electrons was obtained from a high-voltage source by utilizing the Greinacher principle. A separate electron-valve device stabilized the voltage which remained constant within a few hundredths of a percent. The voltage was measured against a standard cell by means of a wire-wound potentiometer. The wave-lengths were obtained from spectra produced by diffraction with etched galena crystals. A beam of electrons defined by two narrow slits fell on the crystal, at grazing incidence, and penetrated small lumps of galena projecting from the surface. The lumps had retained their original orientation to the large crystal and acted as a transmission grating. They gave rise to very sharp spectral lines which could be photographed and measured with high accuracy.

A combination of de Broglie's law with the expression for the Rydberg constant gave e . The most common form of de Broglie's law $\lambda = \frac{h}{mv}$ can be transformed in the following manner:

$$\lambda \sqrt{\left(1 + \frac{eV}{2m_0 c^2}\right)} = \lambda' = \frac{h}{\sqrt{(2em_0 V)}},$$

v is the velocity of the electron, λ the electron wave-length, and V the voltage accelerating the electrons. The square root following λ expresses the influence of velocity on mass. λ' has been introduced to facilitate the calculations.

According to Bohr:
$$R = \frac{2\pi^2 e^4 m_0}{c h^3}.$$

For the charge of the electron we get:

$$e = \frac{V^{\frac{1}{4}} \cdot \lambda'^{\frac{1}{4}}}{\sqrt[4]{e/m_0}} \cdot A, \quad \text{where} \quad A = \sqrt{\frac{R}{c}} \cdot \frac{\sqrt[4]{2}}{\pi} \cdot 10^6.$$

The constants R and c are known with very high accuracy, and e/m_0 which is less well known ($\sim 0.1\%$) enters in the power $1/4$ and thus does not cause an uncertainty exceeding $1/40\%$. The determination gave a value for the electronic charge $e = (4.796 \pm 0.005) \times 10^{-10}$ e.s.u.

The limits of error are meant to take into account all known sources of error.

In the same year Kellström redetermined the viscosity of air, η . This enters into the calculations of the oil-drop value of e in the power $3/2$. Millikan used a value by Harrington $\eta_{23} = 18,227 \times 10^{-8}$. The value had been chosen as being apparently the best of a number of values which varied greatly among themselves.

Therefore, the thought was near at hand that the error in e was due to the use of Harrington's η . Kellstrom's results show that the value $\eta_{23} = 18,227 \times 10^{-8}$ is most probably about 0.67% too low. He got a value $\eta_{23} = 18,349 \times 10^{-8}$ which has been confirmed by Bond who obtained exactly the same result from a different method. This means that Millikan's value of the electronic charge must be increased by about 1% to $(e = 4.818 \pm 0.011) \times 10^{-10}$ e.s.u. The error $\pm 0.011 \times 10^{-10}$ e.s.u. comprises that part only of the total error which comes from uncertainty in η . The appropriate limits of error are even greater.

Finally, Bäcklin and Flemberg in 1936 used the oil-drop method for a new determination. A great improvement was gained by the use of diffusion pump oil of very low vapour pressure. This implies the advantage

that the loss of mass due to evaporation from the droplets has been made imperceptibly small, so that a very great number of observations of the velocity of fall and rise could be made for each individual droplet. Bäcklin and Flemberg (1936) give as the preliminary result of their experiment the value $e = 4.800 \times 10^{-10}$ e.s.u. It has been computed with Kellström's η . The agreement with the X-ray and electron wave values is very satisfactory.

For many years Millikan's value 4.774 has been one of the most familiar numbers to all students of physics but now it is quite obvious that it cannot be maintained any longer.

Table III gives a survey of the values of e obtained by the different methods.

The most satisfactory agreement of the more recent determinations by the X-ray method shows that we know fairly well the ratio between the wave-lengths as measured on the optical and on the conventional X-ray scale. The factor of reduction is $\frac{\lambda_{\text{opt}}}{\lambda_{\text{X-rays}}} = 1.0020$. The value has been verified

by direct comparisons between optical lines and high-order X-ray lines in concave grating spectrographs. There might be, however, some small uncertainty in the calculation of e from this ratio. Du Mond and Bollman (1936) have redetermined, on the one hand, the density of calcite and, on the other, the dimensions of the calcite unit rhombohedron. By using Bearden's wave-length measurements they recalculated e with their new data and got the value $e = 4.799 \times 10^{-10}$ e.s.u.

It has been said against the electron-wave method that the etching of the crystals perhaps could have changed the grating constant of the small crystal lumps exposed by the etching but experimental observations by Du Mond and Bollman and others give no support to such an assumption. They found the same density for very small crystals as for large blocks.

Summing up what has been said above, one finds that the modern determinations give values of the electronic charge which group themselves about the value

$$e = (4.800 \pm 0.005) \times 10^{-10} \text{ e.s.u.,}$$

which ought to give, with its limits of error, rather reliable directions for the search after the true value.

PLANCK'S CONSTANT (h)

Planck's constant can be measured in many different ways but, nevertheless, it is less well known than any other of the four constants discussed in this paper. The experiments never give h alone; it is always combined

TABLE III—ELECTRONIC CHARGE

No.	Method	Experimenter	Year	e	Stated error	Notes	References
1	Oil-drop method	Millikan	1913	4.774×10^{-10}	$\pm 0.009 \times 10^{-10}$	—	(1913)
2	"	"	1917	4.774	± 0.005	—	(1917)
3	"	"	1917	4.770	± 0.005	Recalculated by Millikan (1930)	(1917)
4	"	"	1917	4.818	$\pm 0.011^*$	Recalculated with Kellsiröm's η (1936)	(1917)
5	"	Bäcklin and Flemberg	1936	4.800	—	Kellsiröm's η (1936)	(1936)
6	X-ray method	Bäcklin	1928	4.793	± 0.015	—	(1928)
7	"	Bearden	1931	4.806	± 0.003	—	(1931)
8	"	"	1935	4.8036	± 0.0005	—	(1935)
9	"	Bäcklin	1935	4.805	± 0.004	—	(1935)
10	"	Du Mond and Bollman (Bearden)	1936	4.799	± 0.007	Redetermination of the properties of calcite	(1936)
11	Electron diffraction method	von Friesen	1935	4.797	± 0.005	Recalculated with $e/m = 1.7585$	(1935)

* Error from η only.

with one of the constants e , e/m and m . The combination h/e^n with $n = 3/3$, $4/3$, or $5/3$, occurs most frequently. The uncertainty of this other constant is added to the experimental errors.

Table IV contains those determinations of h , which are of interest to us in this connexion. The Table gives for each determination on the one hand the value published by the experimenter, on the other a value recalculated with more modern values of constants entering the calculations. In some cases incomplete information only is given as to values originally used. For this reason an additional error of a unit or two in the fourth place might have entered some of the recalculated values. When regarding the errors it must be remembered that Feder and Schaitberger do not consider the error of e when stating limits of error.

With the method based on the photoelectric effect, h is determined from the maximum energies which photoelectrons are able to attain at illumination with light of a number of known frequencies. It is difficult to measure this maximum energy accurately and for that reason the value of h obtained is less accurate than those given by the methods II, IV, and V of our Table. Lukirsky's and Priležav's work, in particular, has been subject to severe criticism.

With the X-ray method, associated values of wave-length and voltage are observed near the point where the emission of the continuous X-ray spectrum begins. One of the magnitudes wave-length and voltage is kept constant and the other is varied in small steps. The intensity of the radiation is measured and one extrapolates to zero intensity. It is difficult to know how to make this extrapolation, and it constitutes, therefore, the most serious source of error in this method. It seems necessary to assume an error of at least two-tenths of a percent, if the uncertainty of the value of e is taken into account.

Gnan measured the wave-length of electrons, the velocity of which had been measured according to Kirchner's method of deflexion by alternating electric fields. He found h/m . He measured the wave-length on the crystal scale and, therefore, it is necessary to reduce his value to the absolute scale. The method is of great interest, but it has not been able as yet to give a very reliable value of h .

The other electron diffraction method (von Friesen) is experimentally exactly the same as that described above for the determination of e , the only difference being the elimination of e between the equations instead of h . The author's opinion is that the limits of error may reasonably be taken to cover the true value of Planck's constant.

The last value of Table IV has been derived from the formula for the

TABLE IV—PLANCK'S CONSTANT

No.	Method	Experimenter	Year	Constant entering	h given	h recalculated	Stated error	References
1	I. Photoelectric effect	Millikan	1916	$e^{1/2}$	6.57	6.61	± 0.03	(1916)
2	"	Lukursky and Priležajev	1928	"	6.543	6.58	± 0.014	(1928)
3	"	Olpin	1930	"	6.561	6.60	—	(1930)
4	II. X-ray method	Duane and others	1921	$e^{1/2}$	6.556	6.614	± 0.009	(1921)
5	"	Feder	1929	"	6.547	6.602	± 0.003	(1929)
6	"	Kirkpatrick and Ross	1934	"	6.546	6.602	± 0.006	(1934)
7	"	Schautberger	1935	"	6.555	6.610	± 0.002	(1935)
8	III. Electron diffraction with direct determination of velocity	Gnan	1934	m	—	6.61*	—	(1934)
9	IV. Electron diffraction	von Friesen	1935	$(e/m)^{1/4}$	6.610	6.612†	± 0.012	(1935)
10	V. From Rydberg constant	—	—	$e^{1/2}$	—	6.619	± 0.013	—

* Recalculated to the absolute wave-length scale.

† Recalculated with $e/m = 1.7585$.

Rydberg constant by use of those values of c , e , and e/m which have been chosen above. The accuracy of this value is of about the same magnitude as for the other values.

An inspection of Table V, which gives a survey of the results from different methods, shows that the agreement is quite satisfactory. This fact disproves the statement that the use of Millikan's value of e should

TABLE V—MEAN VALUES OF h

Method	h	Limits of error
Photoelectric effect	6.60	± 0.03
X-rays	6.607	± 0.015
Electron diffraction I	6.61	± 0.02
Electron diffraction II	6.612	± 0.012
Rydberg constant	6.619	± 0.013

alone be able to reconcile the results from the different methods. The limits of error in this Table include the error of the electronic charge. We shall choose as a reasonable value for Planck's constant

$$h = (6.610 \pm 0.015) \times 10^{-27} \text{ erg sec.},$$

and there is reason to believe that the true value will be found inside its limits

MASS OF HYDROGEN ATOM (M_H)

The most reliable method of obtaining M_H is based upon the determination of Avogadro's number N_0 . This number may be calculated from the formula $N_0 = \frac{F \cdot c}{e}$. Its reciprocal equals the mass of an atom of unit atomic weight. If we know, therefore, N_0 and the atomic weight of the atom ^1H we know also the mass of the hydrogen atom. The formula for N_0 contains F , e , and c . Birge has chosen for the Faraday a value 9648.9 e.m.u./g. equiv. In order to reduce this value from the chemical to the physical scale, we must apply a factor 1.00022 obtained from the determination of the relative abundance of the isotopes of oxygen. We get the value $F = 9651.1$ which used with the values $e = 4.800 \times 10^{-10}$ e.s.u. and $c = 2.9978 \times 10^{10}$ cm./sec. gives $N_0 = (6.028 \pm 0.008) \times 10^{23}$.

Aston's new value for the atomic weight of ^1H is 1.00812 and so we find $M_H = (1.673 \pm 0.003) \times 10^{-24}$ g.

The value is higher than Birge's by about two-thirds of a percent. The difference is due chiefly to the use of the higher value for e . e/M_H on the other hand is changed by about one-hundredth of a percent only from 9574.5 (Birge) to 9573.4 e.m.u./g. The ratio M_H/m becomes 1837. In both

N_0 and M_H the dominating source of error is the uncertainty of the value of e .

THE VALUE OF SOMMERFELD'S FINE STRUCTURE CONSTANT

Sommerfeld's fine structure constant $\frac{2\pi e^2}{hc}$ is a pure number According to a hypothesis by Eddington its inverse value is exactly equal to 137. With the constant values chosen above we get 136.9, which, within the limits of error, agrees with the hypothetic value.

RESULTS OF THE SURVEY

Velocity of light	$= (2.9978 \pm 0.0002) \times 10^{10}$ cm /sec.
Specific charge of the electron	$= (1.7585 \pm 0.002) \times 10^7$ e.m.u./g.
Electronic charge	$= (4.800 \pm 0.005) \times 10^{-10}$ e.s.u.
Planck's constant	$= (6.610 \pm 0.015) \times 10^{-27}$ erg. sec.
Avogadro's number	$= (6.028 \pm 0.008) \times 10^{23}$
Mass of hydrogen atom	$= (1.673 \pm 0.003) \times 10^{-24}$ g.

The \pm terms are estimates of reasonable limits of error (not probable errors).

REFERENCES

- Bäcklin 1928 *Uppsala Univ. Årskr*
 — 1935 *Z. Phys.* **93**, 450.
 Bäcklin and Flemberg 1936 *Nature, Lond*, **137**, 635.
 Bearden 1931 *Phys. Rev.* **37**, 1210.
 — 1935 *Phys. Rev.* **48**, 385
 Birge 1929 *Rev. Mod. Phys.* **1**.
 Cornu 1876 *Ann. Obs. Paris*, **13**.
 de Bray, E. J. Gheury 1934 *Nature, Lond.*, **133**, 948.
 Duane and others 1921 *J. Opt. Soc. Amer.* **5**, 213
 Du Mond and Bollman 1936 *Phys. Rev.* **50**, 524.
 Dunnington 1933 *Phys. Rev.* **43**, 404.
 Feder 1929 *Ann. Phys., Lpz.*, **1**, 497.
 Gnan 1934 *Ann. Phys., Lpz.*, **20**, 361.
 Houston 1927 *Phys. Rev.* **30**, 608.
 — 1934 *Phys. Rev.* **46**, 533.
 — 1935 *Zeeman Verhandeligen, The Hague*, **71**
 Mittelstaedt 1929 *Ann. Phys., Lpz.*, **2**, 285
 Kirohner 1932 *Ann. Phys., Lpz*, **12**, 503
 Kirkpatrick and Ross 1934 *Phys. Rev* **45**, 454.
 Lukirsky and Priležnev 1928 *Z. Phys* **49**, 236
 Mercier 1923 *J. Phys. Radium*, **5**, 168.

- Michelson 1879 *Amer. J. Sci.* 18, 390.
— 1885 *Astr. Pap., Wash.*, p. 235.
— 1924 *Astrophys. J.* 60, 258.
— 1926 *Astrophys. J.* 63, 1.
Millikan 1913 *Phys. Rev.* 2, 109.
— 1916 *Phys. Rev.* 7, 355.
— 1917 *Phil. Mag.* 34, 1.
— 1930 *Phys. Rev.* 35, 1231.
Newcomb 1885 *Astr. Pap., Wash.*, p. 112.
Olpin 1930 *Phys. Rev.* 36, 251.
Poase and Pearson 1932 *Astrophys. J.* 82, 28.
Perrotin 1902 *C.R. Acad. Sci., Paris*, 135, 681.
Perry and Chaffee 1930 *Phys. Rev.* 36, 904.
Rosa and Dorsey 1906 *Bull. U.S. Bur. Stand.* 3, 433.
Shano and Spedding 1935 *Phys. Rev.* 47, 33.
von Friesen 1935 *Uppsala Univ. Årsskr.*
Schaibberger 1935 *Ann. Phys., Lpz.*, 24, 84.
-

The Transport Numbers of the Ions in Solutions of Silver Dodecyl Sulphate

BY O. RHYS HOWELL AND HARRY WARNE

The College of Technology, Manchester

(Communicated by Eric K. Rideal, F.R.S.—Received 1 February 1937)

The electrical conductivities of a number of metallic long-chain alkyl sulphates have been measured at different temperatures (Lottermoser and Puschell 1933; Howell and Robinson 1936).

The curve for each salt at each temperature consists of three distinct ranges.*

Over the first range the conductivity falls as a linear function of the square root of the concentration. The behaviour is typical of that of a simple electrolyte and conforms to the Debye-Hückel theory of ionic interaction; the Onsager slope is only a little greater than for a simple salt. It is therefore generally agreed that over this first range the electrolyte is completely

* Compare the curve for silver dodecyl sulphate, fig. 2.

dissociated into the long-chain ion and the simple ion of opposite sign (contra-ion or gegenion).

Over the second range the conductivity falls extremely rapidly to about one-half its initial value. This rapid fall was discovered by McBain for solutions of the soaps and he attributed it to depression of ionization. This was at that time a reasonable conclusion, since owing to excessive hydrolysis of the soaps at great dilutions, measurements were not made at concentrations small enough to discover the initial portion of the curve, which is probably the same for an unhydrolysed soap as for the long-chain alkyl sulphates and other long-chain electrolytes. The depression of ionization postulated by McBain to account for the fall in conductivity over the second range or rather its modern equivalent of ionic interaction really applies to the first range.

Lottermoser and Puschell account for the fall by the formation of aggregates of neutral colloid, but this presupposes the depression of ionization as originally postulated by McBain and is therefore untenable.

Hartley (1935, 1936) has found the same type of curve for cetyl trimethyl ammonium bromide and cetyl pyridinium bromide (Hartley, Collie and Samis 1936) in which the long-chain is cation and for other electrolytes (Molliet and others 1935), and he postulates that micelle formation occurs at the beginning of the second range. Now McBain introduced the conception of the ionic micelle and postulated its formation at the beginning of the third range to account for the subsequent increase in conductivity, the single aggregate with a multiple charge having a greater mobility than the combined mobility of the component ions moving individually. Hartley, however, accounts for the fall in conductivity over the second range by postulating the presence at this stage of micelles which adsorb gegenions from solution, the decrease in conductivity resulting from the removal of gegenions being greater than the increase due to the formation of micelles. Arguments against this view have already been given by Howell and Robinson (1936), and it is now further shown in the sequel that the theory of micelle formation over the second range is not compatible with the observed change in other electrical properties of the solution with increasing concentration.

Howell and Robinson (1936) have shown for a homologous series of metallic alkyl sulphates, that at the point where the linear fall of the first range is abruptly succeeded by the rapid fall of the second, the rotational volume of the long-chain ions is proportional to the free distance between them. It is therefore postulated that at this stage the long-chain ions have been brought to within a critical distance at which their mutual repulsion prevents

them from moving past one another. Under the influence of the electric field, therefore, they move as a complete network and the contra-ions can penetrate in the opposite direction only with difficulty. With increasing concentration the network becomes more densely packed and the movement of the contra-ions is increasingly hindered and eventually stopped, so that over this range the conductivity is greatly reduced. It is shown in the sequel that this view is in complete harmony with the observed changes.

At the end of the second range the sudden fall is arrested, and over the third range the conductivity remains nearly constant; there is, indeed, a slight rise to a maximum followed by a gradual slight fall. It is generally agreed that micelles are present over the third range.

McBain's conception of the ionic micelle and the application of Stokes's law to its movement are sufficient to account for the rise in conductivity provided the micelles have a multiple charge (Howell and Robinson 1936), but it is shown in the sequel that this is not readily compatible with the observed change in the transport number and the mobility of the ions.

Hartley, having postulated the formation of micelles and the adsorption of gegenions on them to account for the fall in conductivity over the second range, was unable to explain the subsequent increase in conductivity over the third range, but he has recently suggested (Hartley and others 1936) that with increasing concentration the adsorbed ions are desorbed again. It is now shown that this mechanism is not compatible with the observed changes in the properties of the ions over the third range.

Howell and Robinson (1936) have shown for a homologous series of the alkyl sulphates and of the soaps, that at the transition from the second to the third range, the distance between the long-chain ions is proportional to their length and that micelle formation then results from the crowding of the long-chain ions to within this critical distance. It is now shown that the subsequent rise in conductivity and the change in the transport number and the mobilities of the ions with increasing concentration can be explained by the formation of new micelles. A rational explanation is thus afforded of the different characteristics found in the two ranges and of the transition from one to the other.

Before the present investigation was undertaken, a few measurements of the transport numbers of the soaps (McBain, Laing and Titley 1919) and of some dyestuffs (Moilliet and others 1935) had been made, but since its completion, two compounds closely related to the alkyl sulphates have been examined, viz. cetyl trimethyl ammonium bromide and cetyl pyridinium bromide (Hartley and others 1936). These are discussed below.

EXPERIMENTAL

Materials—The silver dodecyl sulphate was prepared as follows. Commercial dodecyl alcohol was fractionated and the fraction boiling between 145–149°/15 mm. was accepted. Each batch was estimated by esterifying and analysis by hydrolysis of the ester. The molecular weight thus found varied from 187 to 192, theoretical 186. 1 g. mol. of the alcohol was treated with 1.1 g. mol. of chlorosulphonic acid by gradual addition over 1½ hr. at 40–45°. After ½ hr. further stirring, the mixture was poured into a solution of 1.25 g. mol. of sodium hydroxide in 500 c.c. of water and stirred overnight.

The precipitated sodium dodecyl sulphate was filtered, twice recrystallized from water (40°) and then from methyl alcohol, washed with ether and dried. Analysis *

	Found %	Theoretical %
Hydrolysable SO ₃	27.44	27.70
Ash	24.4	24.65
Inorganic salts	Nil	—

The theoretical amount of silver nitrate plus 20 % excess was added slowly to a solution of the sodium salt at 30°. This and subsequent operations were performed in a dark room in red light owing to the sensitivity to light of the silver salt. The precipitated silver salt was filtered and twice recrystallized from water. It exhibited the very curious behaviour that on drying, even *in vacuo* at atmospheric temperature, it hydrolysed, it could be kept unchanged as the moist solid or in solution. Analysis* showed the salt to be pure.

All solutions were made with specially distilled water of low conductivity.

Electric Circuit—Current was obtained from a battery of accumulators, 120–180 V being used. A Ferranti milliammeter was placed in the circuit to give visual indication of the amount of current passing, but the true amount was obtained from silver coulometers. These consisted of platinum dishes containing 20 % silver nitrate solution with anodes made of flat spirals of thick silver wire covered with filter paper. Two were used, one on the cathode side and the other on the anode side of the transport apparatus, as a check against leakage of current. The weights of silver deposited in the coulometers did not differ by more than 0.0001 g.

Transport Apparatus—The transport apparatus (fig. 1) was a modified

* Analysis kindly carried out by Messrs Imperial Chemical Industries, Dyestuffs Group.

form of the Hittorf H tube; its design is the result of repeated trials and it was found very satisfactory in operation. It is of borosilicate glass and consists essentially of three separately stoppered chambers which can be isolated after passing current by means of the two large-bore taps. The whole contents of each chamber can then be poured out for analysis. The electrodes are flat spirals of thick silver wire welded to platinum which is sealed through the leading tubes, the small piece of platinum in contact with the solution is covered with wax.

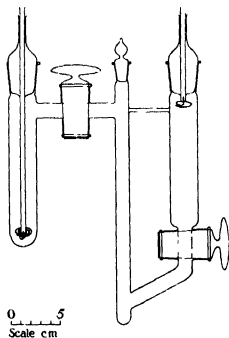


FIG. 1

The vessel was immersed up to the neck in a thermostat adjusted to $40 \pm 0.01^\circ \text{C}$. and the whole was covered to protect the silver salt from the action of light.

Analysis—The silver dodecyl sulphate was handled as a saturated solution at 40°C ., the solutions of lower concentrations being prepared by dilution. Each solution was analysed before and after passage of current. It was boiled with a little acid to hydrolyse the salt and subsequently titrated against standard ammonium thiocyanate solution of approximately the same normality, using saturated iron alum solution as indicator.

RESULTS

The results are given in Table I and plotted in fig. 2. The value of the transport number at infinite dilution has been calculated from the mobilities of the ions at 40° C ; the value 86 for the silver ion is read from a curve obtained by plotting the values at different temperatures (Landolt and Börnstein 1927) and the value of 49 for the dodecyl ion is that of Howell and Robinson from the conductivity of sodium dodecyl sulphate, using the value 69 for the sodium ion obtained in the same way as for the silver ion.

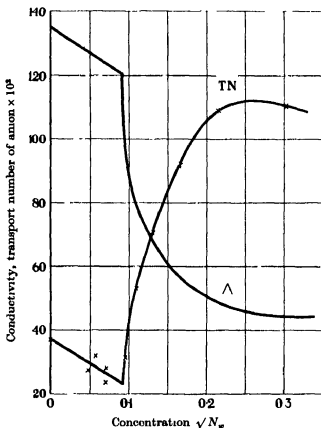


FIG. 2

The first four experimental values are, of necessity, not very accurate. The solutions are extremely dilute and the experimental error in their analysis correspondingly high. That the remaining values are reliable is shown by the good correspondence between the figures obtained from the cathode change and anode change in each determination. It is to be noted

that in dilute solution the cathode value is smaller than the anode value; in concentrated solution, where micelles are present, the reverse obtains. There is a gradual transition doubtless due to a change in the relative hydration of the ions.

TABLE I—SILVER DODECYL SULPHATE. 40° C.

Concentration		Transport number of anion		
N_w	$\sqrt{N_w}$	Cathode value	Anode value	Mean
0	0	—	—	0.363
0.00230	0.0480	0.247	0.289	0.268
0.00334	0.0578	0.286	0.336	0.312
0.00494	0.0703	0.227	0.235	0.231
0.00510	0.0714	0.265	0.286	0.276
0.00655	0.0858	0.302	0.317	0.309
0.0121	0.110	0.511	0.548	0.529
0.0171	0.131	0.704	0.706	0.705
0.0274	0.166	0.917	0.923	0.920
0.0463	0.215	1.124	1.052	1.088
0.0919	0.303	1.193	1.013	1.102

In fig. 2 the conductivity curve for silver dodecyl sulphate is also given. It is drawn from the values of Lottermoser and Puschell, but the initial portion has been amended to conform with the value at infinite dilution due to Howell and Robinson.

The mobilities of the cation and anion have been calculated from the conductivity and transport numbers and they are plotted against the concentration in fig. 3.

DISCUSSION

The curve showing the transport number of the anion against the square root of the concentration of the solution consists of three distinct portions, corresponding with those of the curve for the electrical conductivity (fig. 2). The same correspondence in the change of transport number and of electrical conductivity with concentration is found with other similar substances, e.g. cetyl pyridinium bromide and cetyl trimethyl ammonium bromide (Hartley and others 1936).

First Range—It is generally agreed that over the first range the salt is present as a simple completely dissociated electrolyte. It is seen that the transport number of the anion falls fairly sharply. This is to be expected, since with increasing concentration the enormous anions become relatively far more crowded than the small cations. It is seen (fig. 3) that the mobility of the anion also falls rapidly from 49 to 28, a decrease of 43 %.

The mobility of the cation rises slightly from 86 to 92, an increase of 6.5 %. Normally, in simple salts the mobility of the cation also decreases; for the silver ion the decrease is about 7 % over this range at 18° C. Owing to the great difficulties of measurement in these extreme dilutions, the exact

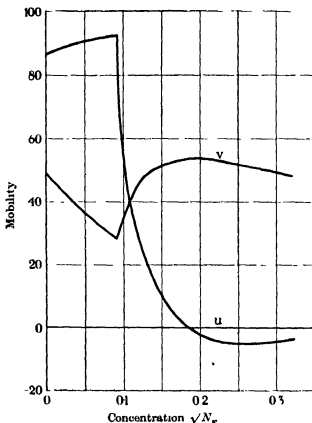


FIG. 3

form of the mobility curves over this first range is uncertain, but if the rise in mobility of the cation is real, it must be due to the special conditions obtaining in the presence of long-chain ions, possibly dehydration resulting in increased mobility.

If the conductivity data of Lottermoser and Puschell be accepted, the mobility of the cation falls by about 14 % over the first range, but that of the anion increases by about 18 % and this is highly improbable. It can hardly be argued that the causes of an increase in mobility of the anion over the second range are already beginning to be operative in the first range, since it is evident from the conductivity curves that there is a very abrupt transition from the first to the second range.

Second Range—Over the second range the transport number of the long-chain anion increases very rapidly to a value exceeding unity. According to the theory already advanced to account for the rapid fall of conductivity over this range (Howell and Robinson 1936), the free space between the long-chain anions having been reduced to within a critical value, they are unable to pass one another and therefore move as a complete network through which the cations are unable to penetrate freely. With increasing concentration the movement of the cations is increasingly hindered; it may eventually be stopped and even reversed. The transport number of the cation therefore falls and may reach zero or even a negative value; that of the long-chain anion rises correspondingly and reaches unity or even greater values.

The change in the mobilities of the ions with increasing concentration over this range (fig. 3) is in accord with this view. The mobility of the cation falls very rapidly to below zero, its movement through the anionic network having been completely checked. The mobility of the anionic network is initially less than that of the individual ions at infinite dilution, but it increases over this range. This is probably due to decreasing viscous drag of the network with increasing concentration due to the enhanced orientation of the component long-chain ions.

At the end of this range the anionic network collapses into micelles. It has already been shown (Howell and Robinson 1936) for a homologous series of the alkyl sulphates and of the soaps that at this point there is a simple proportionality between the length of the long chains and the distance between them.

It has also been shown (Howell and Robinson 1936) that the conductivity at this point may be calculated on the assumption that micelles are formed and that they obey Stokes's law; for sodium dodecyl sulphate there is good agreement between the calculated and observed values if the micelle is assumed to consist of four dodecyl ions with three sodium ions leaving a single nett charge. Similarly for silver dodecyl sulphate, assuming the micelle to consist of four dodecyl ions with three silver ions adsorbed on it,

$$\begin{aligned}
 \Lambda &= \frac{1}{4}(U_{Ag} + V_{micelle}) \\
 &= \frac{1}{4}(U_{Ag} + V_{Do} \times 4^{\frac{1}{2}}) \\
 &= \frac{1}{4}(86 + 49 \times 4^{\frac{1}{2}}) \\
 &= \frac{1}{4}(86 + 123.5) \\
 &= \frac{1}{4}(209.5) &= 52.4 \\
 \text{Observed value} &= 51.0.
 \end{aligned}$$

The agreement is very fair; an Onsager correction to infinite dilution may be necessary but the amount is uncertain. Indeed, it may be argued that since micelles are first formed at this concentration, the dilution is infinite for the salt ionized in this manner.

The transport numbers may be calculated from the same data. Since the micelle consists of four dodecyl ions carrying three sodium ions, its transport number is

$$\begin{aligned} n_{\text{micelle}} &= \frac{V_{\text{micelle}}}{A} \times \frac{7}{4} \\ &= \frac{30.9}{52.4} \times \frac{7}{4} = 1.03 \end{aligned}$$

Observed value = 1.05.

The agreement is satisfactory. Similar data for other compounds are collected in Table II. It is seen that for the soaps both the conductivities

TABLE II

Compound	Temp °C.	Conc. at bend \sqrt{C}	No of chains in mi- celle	Conductivity		Terms for		Transport no of micelle		Remarks
				Calc	Obs.	Contra- ion	Micelle	Calc	Obs.	
Silver dodecyl sulphate	40	0.2	4	52.4	51.0	21.5	30.9	1.03	1.05	
Sodium dodecyl sulphate	20	0.2	4	30.4	30.0	11.4	19.0	1.09	—	
Sodium hexadecyl sulphate	40	0.125	6	33.7	35.0	11.4	22.3	1.21	—	
Potassium laurate	18	0.31	2	48.6	44.0	32.2	16.4	0.51	0.67	At 0.2 N
Potassium oleate	18	0.22	4	29.1	29.5	18.1	13.0	0.78	0.75	At 0.25 N
Sodium oleate	18	0.22	5	20.8	20.5	8.7	12.1	1.06	1.22	At 0.2 N
			6	18.6		7.2	11.4	1.12		
Cetyl trimethyl ammonium bromide	35	0.1	4	40.9	40.0	23.2	17.7	(0.75)	(1.6)	
Cetane sulphonie acid	50	0.16	4	136	c 135	116.0	20.0	(0.26)	(0.56)	

(McBain and others 1919) and the transport numbers* are in good agreement with the calculated values. The conductivities of cetyl trimethyl ammonium bromide and cetyl pyridinium bromide, which are almost identical (Hartley and others 1936), and of cetane sulphonie acid (Hartley, private communication) are also in good agreement with the calculated values, but the transport numbers are much higher than those calculated. Such high transport numbers are compatible with the observed conductivity only if the micelles are very large with multiple charges.

Hartley (1936) holds the view that micelle formation occurs at the

* McBain and Bowden (1923) for potassium laurate and oleate, Laing (1924) for sodium oleate.

beginning of the second range and that the fall in conductivity is due to the adsorption of the gegenions on the micelles. In order to explain the ceasing of the fall and, indeed, the increase in conductivity, which occurs over the third range with increasing concentration, he assumes that the gegenions are desorbed again. This would appear to be improbable (see Hartley and others 1936), and other objections to the theory have already been summarized (Howell and Robinson 1936).

Further illuminating information is obtained by calculating the conductivity of the solution, the transport number of the anion and the mobilities of cation and anion for micelles of different sizes with different charges. This is done in Tables III-VII for silver dodecyl sulphate at 40° using the mobilities of the simple ions at infinite dilution. The changes would, of course, be similar for any other similar electrolyte at any temperature.

TABLE III

Assumed no. of chains in micelle	Assumed no. of charges on micelle	Calculated conductivity of solution	Calculated transport number of anion	Calculated mobility	
				Cation	Anion
4	3	157.1	0.737	41.4	115.7
4	2	104.8	0.885	12.1	92.7
4	1	52.4	1.032	-1.7	54.1

TABLE IV

Assumed no. of chains in micelle	Assumed no. of charges on micelle	Calculated conductivity of solution	Calculated transport number of anion	Calculated mobility	
				Cation	Anion
4	3	157.1	0.737	41.4	115.7
5	2	91.6	1.000	0.0	91.6
6	1	41.3	1.197	-8.1	49.4

TABLE V

Assumed no. of chains in micelle	Assumed no. of charges on micelle	Calculated conductivity of solution	Calculated transport number of anion	Calculated mobility	
				Cation	Anion
4	1	52.4	1.032	-1.7	54.1
5	2	91.6	1.000	0.0	91.6
6	3	123.9	0.979	2.6	121.3

Now it is not clear whether Hartley supposes the whole or bulk of the electrolyte to fall into micelle formation at the beginning of the second range and the adsorption of gegenions to follow with increasing concentra-

tion. If so, the suggested mechanism fails to account for the facts. If the size of the micelles remains constant whilst the charge decreases (owing to absorption of contra-ions), then it is seen from Table III that the conductivity should decrease, the transport number of the anion increase and the mobility of the cation decrease as is found experimentally; but the mobility of the anion should decrease and this is contrary to experimental fact. Similarly, it is seen from Table IV that if the size of the micelle increases whilst the charge simultaneously decreases, the same effect should be found.

TABLE VI

Assumed no. of chains in micelle	Assumed no. of charges on micelle	Calculated conductivity of solution	Calculated transport number of anion	Calculated mobility	
				Cation	Anion
4	2	104.8	0.885	12.1	92.7
5	2	91.6	1.000	0.0	91.6
6	2	82.6	1.088	-7.3	89.8

TABLE VII

Assumed no. of chains in micelle	Assumed no. of charges on micelle	Calculated conductivity of solution	Calculated transport number of anion	Calculated mobility	
				Cation	Anion
4	1	52.4	1.032	-1.7	54.1
3	1	62.6	0.903	5.9	56.7
2	1	81.9	0.712	23.6	58.3

If, on the other hand, it is assumed that micelles are gradually formed over the range, then they cannot be initially highly charged, and the fall in conductivity and change in other properties cannot be explained by the adsorption of contra-ions. It has already been shown (Table II) that at the end of the second range, the conductivity accords with the existence of singly charged micelles composed of four long-chains, and it is seen from Table III that the formation of a micelle of this size with even three charges only would result in increasing conductivity; with two charges the conductivity is little less than that found at the beginning of the range. Evidently, therefore, if micelles are formed over this range, they are not highly charged and the decrease in conductivity cannot be ascribed to the adsorption of contra-ions; the original micelles must contain adsorbed contra-ions leaving a low (probably single) charge. The difficulty would remain of accounting for the difference between the second and third ranges and for the relatively sharp transition from one to the other if micelles are present in both. The three distinct ranges shown in the change of conductivity

and other physical properties with increasing concentration indicate very definitely that the electrolyte is present in three different states.

Third Range—Over the third range it is generally agreed that micelles are present as originally postulated by McBain. It seems to have been implicitly assumed that with increasing concentration, the micelles increase in size.

Table V shows the change in the conductivity, transport number and mobilities calculated for silver dodecyl sulphate at 40° C. on the assumption that the micelle and its charge increase together. It is seen that the conductivity should increase, the transport number of the anion decrease and the mobility of the cation increase (but only very slightly), all these changes being in accord with the experimental findings, the mobility of the anion, however, should increase very rapidly, whereas it is found experimentally to decrease.

It is evident from Table III (read in reverse order) that if the size of the micelle remains constant and the charge increases the same changes should occur in a still more marked degree. Obviously, therefore, Hartley's suggestion that over the third range the gegenions are desorbed, fails to account for the observed facts.

If it is assumed that the size of the micelle increases while its charge remains constant (Table VI) the conductivity should decrease, the transport number of the anion increase and the mobility of the cation decrease, all of these changes being in opposition to the experimental facts.

Now when micelle formation first occurs at the beginning of the third range, the space between the ions is thereby greatly increased. The distance between the micelles will be greater than that previously existing between the long-chain ions, and the number of contra-ions in a given space will be reduced to one quarter its former value. It is therefore reasonable to suppose that with further increasing concentration of electrolyte new micelles will be formed and that these may initially be smaller than those already existing.

It is seen (Table VII) that this would result in increasing conductivity, decreasing transport number of the anion and increasing mobility of the cation, all of which are in accord with experimental observation. There would be an increase in the mobility of the anion but it is very small (only 4 units as against the decrease of 25 units for the cation). The calculated values have to be corrected for the interionic reaction due to increasing concentration. The increase in conductivity, the decrease in transport number of the anion and the increase in mobility of the cation will thereby be reduced; the small increase in the mobility of the anion will also be

reduced and converted into a small decrease whilst that of the cation (being so much larger), though reduced, would remain. All the observed facts are therefore explained.

We are indebted to Messrs Imperial Chemical Industries for the gift of material and for permission to publish this paper. One of us (H. W.) also desires to acknowledge receipt of a grant.

SUMMARY

The transport numbers of the ions in solutions of different concentrations of silver dodecyl sulphate have been determined by the Hittorf method at 40° C.

The change in the transport numbers with increasing concentration occurs in three distinct ranges corresponding with those of the electrical conductivity.

Over the first range, where the electrical conductivity falls in accordance with the Onsager equation, the transport number of the anion falls rapidly; the mobility of the anion falls rapidly whilst that of the cation increases slightly. All these facts are consistent with the theory, generally agreed, that the salt is present as a simple completely dissociated electrolyte.

Over the second range, where the electrical conductivity falls extremely rapidly, the transport number of the anion rises rapidly to a value greater than unity; the mobility of the anion increases whilst that of the cation falls rapidly. All these facts are in agreement with theory advanced by Howell and Robinson that over this range, the long-chain ions are unable to pass one another and therefore move as a complete network through which the cations are unable to pass. The alternative theory that micelles are being formed over this range does not account for all the facts.

Over the third range, where the electrical conductivity remains almost constant or increases slightly, the transport number of the anion decreases slightly; the mobility of the anion decreases slightly whilst that of the cation increases slightly. All these facts are compatible with the theory originally postulated by McBain that micelles are present over this range. A full explanation is given by the theory that at the beginning of this range, the long-chain network collapses into micelles and that with further increasing concentration of the electrolyte, new and smaller micelles are produced. It is shown that on the theory of micelle formation over the second range, it is impossible to account for the transition to the different behaviour of the third range.

REFERENCES

- Hartley, G. S. 1935 *Trans. Faraday Soc.* **31**, 31-50.
— 1936 "Aqueous Solutions of Paraffin-Chain Salts." *Actualités Sci. et Industr.*, No. 387.
Hartley, G. S., Collie, B. and Samis, C. S. 1936 *Trans. Faraday Soc.* **32**, 795-815.
Howell, O. R. and Robinson, H. G. B. 1936 *Proc. Roy. Soc. A*, **155**, 386-406.
Laing, M. E. 1924 *J. Phys. Chem.* **28**, 673-705.
Landolt and Bornstein 1927 "Tabellen," 5th ed. 1st supp. vol. Berlin: Julius Springer.
Lottermoser, A. and Puschell, F. 1933 *Kolloidzechr.* **63**, 175-92.
McBain, J. W. and Bowden, R. C. 1923 *J. Chem. Soc.* **123**, 2417-30.
McBain, J. W., Laing, M. E. and Titley, A. F. 1919 *J. Chem. Soc.* **115**, 1279-1300.
Moilliet, J. L., Collie, B., Robinson, C. and Hartley, G. S. 1935 *Trans. Faraday Soc.* **31**, 120-9.
-

The Nuclear β -Rays of Radium D

BY H. O. W. RICHARDSON, *University of Liverpool*
AND ALICE LEIGH-SMITH, *King's College, London*

(Communicated by J. Chadwick, F.R.S.—Received 6 February 1937)

[Plates 2, 3]

The radium D.E. disintegration is of special interest because of the very low energy of the disintegration electrons. It appears probable that more than half of these are emitted with less than 4 kV energy so that the energy change is less than in any other known nuclear disintegration.

While there has been much uncertainty about the energies of the nuclear electrons, certain other facts about the transformation appear to be well established. In nearly every disintegration the product nucleus of radium E is left excited 47 kV above the ground-level. The resulting γ -ray is very strongly internally converted in the *L* and outer atomic levels so that only about 3.5 γ -quanta escape from the parent atom in 100 disintegrations. These results have been established by expansion chamber observations of the β -rays and by observations of the effects produced by the unconverted 47 kV γ -quanta (Bramson 1930; Stahel and Sizoo 1930) and the *L* charac-

teristic X-radiation which is emitted as a consequence of the ejection of a secondary β -ray from the L level by the internal conversion (von Droste 1933; Stahel 1935).

The presence of the copious secondary β -rays with ranges extending from about 1.0 up to 5.0 cm. has tended to mask the much shorter tracks due to the nuclear rays and to superpose on them a continuous spectrum due to reflected secondary electrons, which have been scattered back with varying losses of energy from the material on which the source of radium D is deposited.

In all the expansion chamber measurements (Kikuchi 1927; Petrova 1929; Feather 1929; Richardson 1931) a peak appeared in the distribution curve at a range of about 1 cm. even when reflexion was reduced by mounting the radium D on thin gold leaf. On the basis of identifying this peak with the nuclear electrons, one of us suggested that the nuclear continuous spectrum extended as far as 36 kV (Richardson 1931).

This result did not accord with the counter observations of Stahel (1931). Stahel counted the number of β -rays which passed through a collodion window of 1 mm. stopping power and found only 0.83 particles per disintegration; whereas it appears that the photoelectrons due to the 47 kV γ -ray should give, under these conditions, about 0.96 particles per disintegration, and also that a large fraction of the L -tertiary electrons, which amount to 0.30–0.36 per disintegration should also be observed.

In view of this uncertainty it was felt desirable to attack the problem by a new method in which the occurrence of reflexion was impossible. The solid mounting of the source was eliminated by using the radium D in the form of a vapour so that tracks could be photographed starting in the gas in the expansion chamber (Richardson and Leigh-Smith 1934).

EXPERIMENTAL DETAILS

Lead was mixed with a little of its isotope, radium D, and converted by the Grignard method into the poisonous tetramethyl which has a vapour pressure nearly equal to that of water. A few drops of the ether solution were dropped into a small pot inside the cloud chamber, and after evaporation had occurred the characteristic short tracks of radium D were observed starting in the gas throughout the chamber. The latter had a cylindrical piston lubricated with oil, and alcohol was used as condensant. In 75 expansions the chamber was filled with air and in 54 with mixtures of H_2 and CO_2 of stopping powers varying from 0.24 to 0.40 relative to oxygen at 760 mm. and $0^\circ C$. CO_2 was used to reduce the mobility of the ions

without introducing an oxidizing agent. All the ranges to be mentioned are reduced to the scale of the experimental ranges in the air expansions in air of stopping power 0.75 relative to O_2 . The photographs were taken with a double camera with the lens axes inclined at about 35° to the piston axis. The ranges of the tracks were measured by stereoscopic reprojection, the life-size three-dimensional real image of a track being mapped out by pins, the separation of whose points was then found.

It was found necessary to clean out the chamber a few hours after introducing radium D in order to reduce a source of error due to the growth of the subsequent body, radium E. This cleaning was facilitated by covering the piston with disks of black paper which could be easily renewed, but it considerably reduced the rate of taking photographs. It was necessary to reduce the amount of radium E present on the surfaces of the chamber because the resulting fast β -radiation (1) obscured the photographs and (2) produced secondary branch tracks which under certain conditions could be mistaken for radium D tracks starting in the gas. This is particularly possible if the fast weakly ionizing parent track is travelling nearly along the lens axes of the camera, for then it may easily be invisible, so that the denser branch track will appear to be unaccompanied. This error is more likely (a) when there is much general cloud in the chamber, and (b) when the track is old and diffuse so that the few drops of the fast parent track may be masked by the general background. Further, as the number of branches is roughly inversely as the square of their energy, this source of error is larger for shorter tracks.

THE IDENTIFICATION OF THE SECONDARY ELECTRONS

It was found that the tracks could be classed into two groups, "isolated short" tracks and "long" tracks, most of the latter originating at the same point as one or two very short companion tracks. The range distribution of both the "isolated short" and of the "long" tracks is shown in fig. 1, curve 1 coming from air expansions and curve 2 from expansions in gas mixtures of lower stopping power and therefore greater sensitivity to low energy rays. Both curves 1 and 2 show a well-marked maximum at about 2.0 cm. which has the shape to be expected for a group formed from β -rays of equal energies, the ranges falling on the well-known straggling curve. The energy of the group when calculated from the experimental mean range, \bar{R} , of 1.95 cm., by the formula $E = 24.6(S\bar{R})^{1/1.83}$, where S is the stopping power relative to O_2 at $0^\circ C.$ and 760 mm., comes to 30.3 kV

in good agreement with that of the very strong group known to come from the L_1 level, at 31 kV.

Similarly the tracks above 2.9 cm. may be accounted for as secondaries ejected from the M , N , or outer levels with energies 43.3, 46.1 kV and mean ranges 3.72 and 4.18 cm. The "isolated short" group of seven tracks

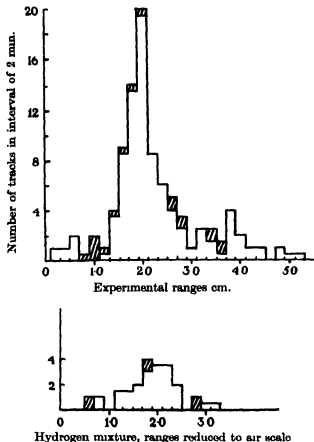


FIG. 1.—Curve 1 expansions in air; curve 2 expansions in hydrogen mixture.

cannot be accounted for in this way but has many possible origins. (a) Isolated nuclear tracks are to be expected in the 3.5 % of disintegrations in which the γ -ray escapes unconverted. (b) The quanta of characteristic L -radiation of bismuth, which according to Stahel (1935) are emitted in 25 % of the disintegrations, will eject photoelectrons from the gas with energies varying from 9.4 to 16.2 kV and ranges extending owing to straggling from 1.5 to 9 mm. (c) The 47 kV γ -quanta will give rise to Compton recoil electrons of ranges below 2 mm. (d) The short track may really be

a radium E branch as discussed above. It is probable that all the sharp tracks below 1.0 cm. range can be explained on the basis of (b). No attempt was made to identify and count isolated tracks shorter than 2 mm.

The tracks marked with cross-hatching (fig. 1) in the distribution are all old diffuse tracks in which only a single layer of either positively or negatively charged drops is visible, the other layer having been drawn by the vertical electric field out of the registering region of the chamber. These old tracks can be seen to have a more random distribution than the others which are either sharp or in eleven cases consist of two layers close together (as in fig. 5*d*, Plate 3). It is plausible that some of these old tracks have started from atoms on the roof or the piston and may have lost energy in escaping, while others may be branches arising from the process (d) above. The whole group is therefore rejected with the result that the ordinate of the distribution at 1.0 cm. is zero, whereas in all previous expansion-chamber observations the distributions have shown a maximum near 1.0 cm. This maximum must have arisen from the reflexion, with loss of energy, of secondary electrons from the mounting material.

THE DETERMINATION OF THE POINTS OF ORIGIN

Most of the long secondary tracks either formed a pair with another very short track, or what may conveniently be termed a trio with two clearly separable very short tracks starting at the same point. In the case of many of the pairs it was found difficult to decide the point of origin, largely owing to the great nuclear scattering of slow β -rays which was markedly increased by the presence of Pb nuclei in the gas, so that many deflexions through more than 140° were found.

It may be remarked that even if no attempt were made to determine the origins and if the ranges in fig. 1 were the sum of the ranges of both components of a pair, no appreciable change in the distribution would be produced because the short component rarely exceeded 2 mm.

By using the following criteria the origin could be determined very closely in most cases:

(a) A magnetic field of 650 gauss curved the tracks. On each side of a nuclear deflexion the curvatures would have the same sense whereas on each side of an origin the curvatures would be opposite. This was particularly useful in looking for the junction of a secondary and a long nuclear track of comparable range, a type of pair which was not found.

(b) The linear ionization along a slow track is roughly inversely as the energy so that a sudden change in the denseness of the track occurs at the

junction where the secondary has perhaps five times the energy of the short track. In some cases, as in fig. 4a, Plate 2, this gave the tracks a curious double-ended appearance.

(c) In the case of well-marked trios there was no uncertainty, the probability that one short component was a branch being small.

THE RANGES OF THE NUCLEAR TRACKS

The occurrence of trios must be due to the emission of a nuclear, a secondary and a tertiary electron from the same atom, the latter electron arising from the internal conversion of a quantum of characteristic X-radiation, the Auger effect (Auger 1926). The tertiary electrons due to the M X-radiation will have too little energy to form tracks exceeding 0.05 cm range so that no trios should be found associated with M or outer secondary electrons. The short track of such a secondary must therefore be nuclear if longer than 0.05 cm. The nuclear range distribution obtained from the 16 disintegrations falling in this group, called class I, is shown in fig. 2. About half the ranges were below 0.05 cm and were too short to measure

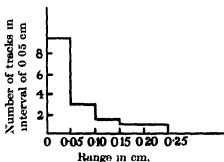


FIG. 2. Ranges of 16 nuclear tracks in class I

It appears from the measurements of Lay (1934) on the L fluorescence efficiency of Pb, that 60% of the L secondaries should be accompanied by tertiaries which cannot be distinguished from the nuclear tracks except in cases where there is no short track exceeding 0.05 cm., when it is certain that the tertiary electron is absent. This is because the tertiary electrons due to the L -radiation form a line spectrum extending from 7 to 12.7 kV and thus may have ranges from above 0.07 to about 0.6 cm. The L disintegrations with no tertiary form class II in Table I and the two further classes are, III, well-marked trios in which both short tracks exceed 0.05 cm and, IV, cases in which one short track only exceeds this range while traces of another shorter track may or may not be visible.

Owing to the overlapping of the nuclear and tertiary ranges it is not possible to obtain a detailed nuclear range distribution from classes II, III and IV. It is however possible to divide the nuclear tracks into two

groups having ranges above and below 0.05 cm. assuming 60 % for the tertiary electron emission. Let n be the fraction of nuclear electrons below 0.05 cm. range. Then the fraction of all the disintegrations with L secondaries (i.e. of classes II, III and IV) falling in class II will be $0.40n$ because 0.40 is the independent probability that no tertiary track is present. Similar expressions hold for III and IV and are given in Table II.

TABLE I—TYPES OF DISINTEGRATION

Class	Level of origin of secondary electron	Range limits of short components	Possible numbers of distinguishable components	Observed number in class
I	M or outer	None	2 or 1	16
II	L	All < 0.05 cm.	2 or 1	24
III	L	Two > 0.05 cm.	3	23
IV	L	One > 0.05 cm.	3 or 2	44

TABLE II

1	2	3	4	5	6
Class	Dependence on n	Observed fractions	Values of n calculated from 3 and 2	"Adjusted" fractions	n values from 2 and 5
II	$0.40n$	0.264	0.660	0.231	0.577
III	$0.60(1-n)$	0.253	0.573	0.253	0.573
IV	$0.40 + 0.20n$	0.483	0.401	0.517	0.559

It can be seen that the values of n in column 4 deduced from the two preceding columns do not agree and that the value of n varies rapidly with the fraction of tracks in class IV. A transfer of 3.5 % of disintegrations from classes II–IV as in column 5 produces agreement among the n values as shown in column 6.

The lack of agreement in 4 may be due to statistical fluctuations in the small numbers involved or more probably to the difficulty of applying the criterion of 0.05 cm., a length near the limit of measurement in the image. Further, the short tracks tended to spread vertically in the electric field owing to local deficiencies of supersaturation and in many cases their origins were not clear owing to bursts of secondary ions along the long track of the pair. It is concluded that 57 % of the nuclear electrons have ranges below 0.05 cm. corresponding to energies below 4 kV with a probable error of about ± 8 %. This accords with the distribution in class I in which 60 % fall below 0.05 cm. range.

COMPARISON WITH THEORY

The distribution curve predicted by the theory of Fermi (1934) with E_0 at 15.75 kV is shown in fig. 3, curve 1, using the form

$$NdE = A(E + 511)(1 + 0.355\eta)(E_0 - E)^2 dE,$$

where N is the number of electrons emitted with energy between E and $E + dE$ in kV, E_0 is the maximum energy and η is the electron momentum in units of mc or $H\rho/1700$ gauss cm. In the Fermi theory the interaction

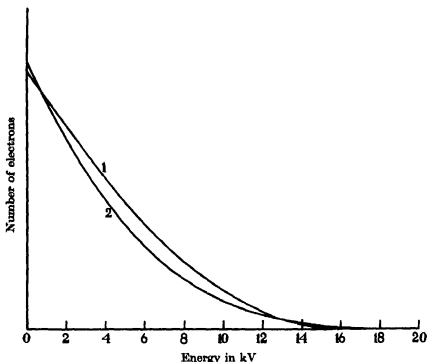


FIG. 3—Curve 1 (0, 0) interaction, $E_0 = 15.75$ kV; curve 2 (0, 1) interaction, $E_0 = 22.5$ kV.

energy between the heavy and light particles depends on the product of the electron and neutrino wave functions at the nucleus. The second curve is the prediction of the modified or (0, 1) interaction proposed by Konopinski and Uhlenbeck (1935) in which the first derivative of the neutrino wave function is introduced in the interaction energy. The expression for NdE is the same as before except that the power of $(E_0 - E)$ is raised from 2 to 4 corresponding to a bias in favour of high neutrino energies.

Taking E_0 now as 22.5 kV the curve 2 in fig. 3 is obtained giving 61.6% of the electrons below 4 kV in rough agreement with the experimental

divinoning. Raising E_0 to 30.3 kV changes this to 49 % below 4 kV but gives about 65 % above 13 kV in comparison with 1.4 % predicted by $E_0 = 22.5$ kV

13 kV corresponds to a mean range of 4.17 mm and only three short tracks were found which exceeded this of ranges 6.5 and 4.5 mm. In each case they were associated with L secondaries and it is therefore impossible to distinguish them from tertiary electrons which would occasionally have ranges as long as this. The experiments therefore favour the lower value of E_0 . They would be equally satisfied by a (0,0) or Fermi distribution formula with k_0 at 16 kV such as shown by curve 1 fig. 3.

Both the (0,0) and (0,1) formulae predict a maximum ordinate at $E = 0$ and that about 10 % of the nuclear electrons should be below 0.5 kV corresponding to as little as about 15 ion pairs in the track. Four of these low energy disintegrations should appear among the 24 tracks in class II. In all but three of these 24 tracks an initial burst of ions is visible and in two of the three cases the track is diffuse so that presence of an initial burst may be masked by the ordinary linear ionization of the secondary averaging about 25 ion pairs per mm. Further about 70 % of the I secondaries should be accompanied by quaternary electrons having energies of about 3 kV which would be indistinguishable from the shorter nuclear tracks. These quaternary electrons discussed by Auger (1926) arise from the internal conversion of M X radiation following the replacement by an M electron of the secondary ejected from the I level. These tracks will lie below the 0.05 cm limit but they make it impossible to assert that the initial bursts of ions observed in classes I and II are entirely nuclear in origin.

For such low energies the effect of screening by the atomic electrons may be appreciable. As considered by Rose (1936) the effect can be divided into (a) a general raising of the energy of the electron levels near the nucleus by about 11 kV above their energy in the absence of the atomic electrons and (b) an interference between the escaping electron and the outer electrons giving rise to a reflexion at the screening potential barrier.

The effect (a) would raise E_0 by 11 kV above its value for a bare nucleus and would cause all electrons below 11 kV to escape only through this effective lowering of the potential barrier. The effect (b) might slightly reduce the ordinate at low energies.

THE β TRANSITION BETWEEN THE GROUND STATES

Another point is the possible occurrence of long range nuclear electrons corresponding to a value of E_0 raised by 47 kV the energy of the excited state of radium E in accordance with the theory of Ellis and Mott (1933)

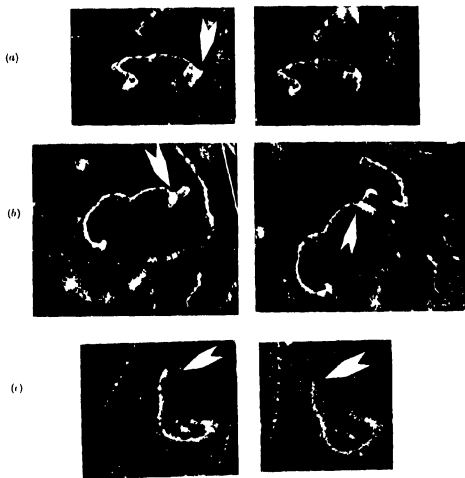


FIG. 4.—*L* track in air. Ranges 1.7, 0.2, and about 0.1 cm. (b) *L* track in air. Ranges 2.35, 0.3, and about 0.1 m. (c) *L* secondary in air. Range 2.05 cm. No nuclear track visible.

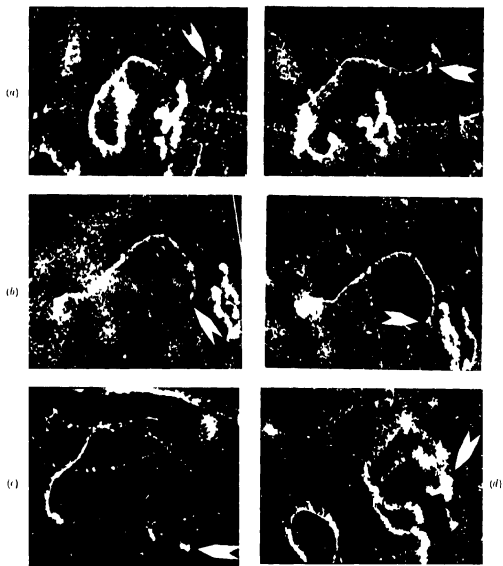


FIG. 5.—(a) *M* or *N* secondary in air. Range 1.35 cm. Nuclear track about 0.2 cm. (b) *M* secondary in air. Range 3.2 cm. No nuclear track visible. (c) *L* pair in hydrogen mixture. Reduced ranges 2.4 and 0.1 cm. (d) *L* trio in hydrogen mixture. Reduced ranges 1.8, 0.4, 0.27 cm. The tracks are old and diffuse and the ions have separated into a positive and negative layer.

Such electrons would be unaccompanied by secondaries and would have ranges extending up to about 8 cm. Taking the (0, 1) formula with $E_0 = 72.2$ kV, 36 % of these electrons will have energies between 10 and 25 kV and will correspond to a part of the range distribution in which the experimental ordinate is in fact very small. Almost all the tracks above 1.4 cm. range which are not pairs or trios are accounted for in classes I and II, and all except three in these classes show traces of an initial "spur", which should be absent on the long-range hypothesis. The few tracks below 1.4 cm. can be explained as photoelectrons ejected from the gas. It is concluded that the β -ray transition between the ground states of radium D and E cannot take place in more than about 10 % of the disintegrations and therefore corresponds to a forbidden β -transition.

SUMMARY

The nuclear electrons of radium D have been studied by photographing the tracks of the β -rays from a radioactive vapour of radium D in a cloud chamber.

As a result of measuring the β -ray tracks emitted in 107 radium D disintegrations it is concluded that 57 ± 8 % of the nuclear electrons have energies below 4 kV and that not more than about 3 % can have energies exceeding 13 kV. This is consistent with a Konopinski-Uhlenbeck end-point at about 24 kV or a Fermi end-point of about 16 kV. No evidence for the presence of a "long-range" β -transition between the ground states of radium D and E was found and, if present, it must occur in less than about 10 % of the disintegrations.

REFERENCES

- Auger, P. 1926 *Ann. Phys. Radium*, **6**, 183-253.
 Brainson, S. 1930 *Z. Phys.* **66**, 721-40.
 von Droste, G. Frh. 1933 *Z. Phys.* **84**, 17-41.
 Ellis, C. D. and Mott, N. F. 1933 *Proc. Roy. Soc. A*, **141**, 502-11.
 Feather, N. 1929 *Proc. Camb. Phil. Soc.* **25**, 522-9.
 Fermi, E. 1934 *Z. Phys.* **88**, 161-77.
 Kikuchi, S. 1927 *Jap. J. Phys.* **4**, 143-58.
 Konopinski, E. J. and Uhlenbeck, G. E. 1935 *Phys. Rev.* **48**, 7-12.
 Lay, H. 1934 *Z. Phys.* **91**, 533-50.
 Petrova, J. 1929 *Z. Phys.* **55**, 627-45.
 Richardson, H. O. W. 1931 *Proc. Roy. Soc. A*, **133**, 367-80.
 Richardson, H. O. W. and Leigh-Smith, A. 1934 *Nature, Lond.*, **134**, 772-3.
 Rose, M. E. 1936 *Phys. Rev.* **49**, 727-9.
 Stahel, E. 1931 *Z. Phys.* **68**, 1-11.
 — 1935 *Helv. Phys. Acta*, **8**, 651.
 Stahel, E. and Sizoo, G. J. 1930 *Z. Phys.* **66**, 721-40.

The Adsorption of Gases and the Equation of the Liquid State

By E. C. C. Baly, F.R.S.

(Received 23 January 1937)

It has been shown in previous communications (Baly 1935, 1936) that alumina and NiO are adsorbed by kieselguhr and that the adsorbed layer is activated. In the case of alumina only a unimolecular layer is adsorbed, but in the case of NiO three layers are progressively adsorbed.

The mechanism of the activation of the adsorbed layers is in all probability the same as that already established in the case of solvates (Baly 1928). If the molecules in the surface are denoted by S and those of the adsorbate by A the adsorption complex may be denoted by $S-A^+$, namely a complex in which the molecule S has donated one or more of its rotation or rotation-vibration quanta to the molecule A . The complex is stabilized by the donation of this energy and will be resolved into its components by the supply to it of an amount of energy equal to that donated by the surface molecule.

The donation of energy in the formation of these complexes depends on the two components, surface molecules and adsorbate molecules, possessing similar frequencies in the infra-red. It is not necessary that the two frequencies be identical. It is known that, except at temperatures of the order of -280°C , the rotation-vibration bands comprise a number of associated frequencies, and it is only necessary that the bands of the two components overlap. A somewhat similar phenomenon is observed in fluorescence, where the whole of the associated frequencies in the fluorescence band are excited by any single frequency characteristic of the relevant absorption band.

Let the case be considered of a gas brought into contact with a clean solid surface. If S be the number of molecules in that surface and A the molecular concentration of the gas, the initial velocity of the formation of the adsorption complexes will be k_1AS and, if the volume of the gas be large and the pressure maintained constant, the velocity at time t will be $k_1A(S-p)$, where p is the number of adsorption complexes which have been formed. The velocity of decomposition of the adsorption complexes at that time will be $k_2pe^{-Q_1/RT}$, where Q_1 is the amount of energy donated by a molecule S to a

molecule A in forming the adsorption complex. It is evident that in a closed system an equilibrium state will be reached which is defined by

$$k_1 A(S-p) = k_2 p e^{-Q_1/RT}, \quad (1)$$

where A is the molecular concentration of the gas at equilibrium. From this expression we have

$$p = \frac{k_1 A S}{k_2 e^{-Q_1/RT} + k_1 A} = \frac{A S \frac{k_1}{k_2} e^{Q_1/RT}}{1 + A \frac{k_1}{k_2} e^{Q_1/RT}}$$

and

$$\frac{p}{S} = \frac{A \frac{k_1}{k_2} e^{Q_1/RT}}{1 + A \frac{k_1}{k_2} e^{Q_1/RT}} = \frac{aA}{1 + aA}, \quad (2)$$

where

$$a = \frac{k_1}{k_2} e^{Q_1/RT}.$$

Since p/S is the fraction of the surface covered by the adsorption complexes, equation (2) is identical in form with the well-known Langmuir expression and has the advantage of directly indicating the effect of temperature change.

Now there is a great deal of evidence that the adsorption of gases is by no means restricted to a single molecular layer, and it is possible to calculate by the foregoing method the number of molecules absorbed in the second, third, etc., layers. If the number of gas molecules in the first adsorbed layer be p , the initial rate of formation of the second layer of adsorption complexes will be $k_3 A p$, and after time t the rate will be $k_3 A(p-q)$. The rate of dissociation of q complexes in the second layer will be $k_4 q e^{-Q_2/RT}$, where Q_2 is the energy donated by a molecule in the first adsorbed layer to a molecule in the second layer. The equilibrium state will be defined by

$$k_3 A(p-q) = k_4 q e^{-Q_2/RT},$$

whence

$$q = \frac{A p \frac{k_3}{k_4} e^{Q_2/RT}}{1 + A \frac{k_3}{k_4} e^{Q_2/RT}} = \frac{b p A}{1 + b A}, \quad (3)$$

where

$$b = \frac{k_3}{k_4} e^{Q_2/RT}.$$

Substituting the value of p from (2) we have

$$q = S \left[\frac{aA}{1 + aA} \times \frac{bA}{1 + bA} \right]. \quad (4)$$

Similarly the number of molecules in the third adsorbed layer will be given by

$$r = S \left[\frac{aA}{1+aA} \times \frac{bA}{1+bA} \times \frac{cA}{1+cA} \right], \quad (5)$$

where

$$c = \frac{k_5}{k_6} e^{Q_5/RT},$$

and in the same way the number of molecules adsorbed in the fourth, fifth, etc., layers may be obtained. The total number of molecules adsorbed on the surface will be given by

$$\alpha = S \left[\frac{aA}{1+aA} + \left(\frac{aA}{1+aA} \times \frac{bA}{1+bA} \right) + \left(\frac{aA}{1+aA} \times \frac{bA}{1+bA} \times \frac{cA}{1+cA} \right) + \dots \right] \quad (6)$$

The definition of the adsorption complex as being formed between a molecule in the surface of the adsorbent and a molecule of the adsorbate is not all-embracing. If the molecules of the adsorbate are large compared with the molecules in the surface, the adsorption complex will be formed between one molecule of the gas and more than one molecule in the surface. Then again in the case of metallic surfaces it is not possible to speak of molecules. The correct definition of S is an adsorbing unit which may consist of one or more atoms or molecules.

Before discussing the application of the adsorption formula (6), it is advisable to determine the type of adsorption curve it gives, and for this purpose it is necessary to assign values to the various exponential terms $\frac{k_1}{k_2} e^{Q_1/RT}$, $\frac{k_3}{k_4} e^{Q_3/RT}$, etc. In so doing it may be accepted without argument that the first adsorbed layer is much more stable than the second, that is to say $\frac{k_1}{k_2} e^{Q_1/RT}$ is much greater than $\frac{k_3}{k_4} e^{Q_3/RT}$. Furthermore, there can be no doubt that the exponential terms progressively decrease for each consecutive layer after the second. In order to illustrate the relation between the amount of gas adsorbed and the equilibrium pressure, the following series of exponential terms were taken, 4×10^8 , 4×10^5 , 4×10^4 , 8×10^3 , 2×10^3 , 6×10^2 , 3×10^2 and 2×10^2 , and it was assumed that the temperature was above the critical temperature of the gas.

The relation between the values of $\log \alpha/S$ and $\log A$ calculated from this formula is shown by the curve shown in fig. 1, and this curve is of considerable interest. As indicated it may be divided into three portions, namely, that between $\log A = 9.000$ and 7.778 , that between $\log A = 7.778$ and 4.301 and

that between $\log A = 4.301$ and 1.000 . The central portion, though in fact sigmoid, can be expressed by the linear equation

$$\log \alpha/S = 1.267165 + 0.1917913 \log A,$$

that is to say

$$\alpha/S = 18.49971 A^{0.1917913}.$$

The values of α/S calculated by this formula are given in the second column of Table I, and the agreement with the values obtained from the adsorption formula and given in the third column is remarkably good.

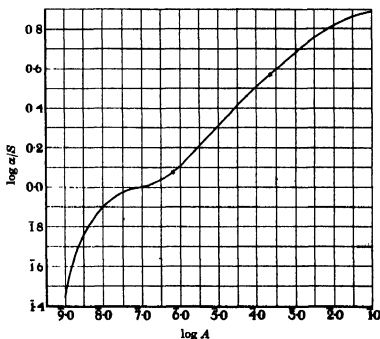


FIG. 1

TABLE I

A	α/S (calc.)	α/S	A	α/S (calc.)	α/S
4×10^{-7}	1.0987	1.1330	2×10^{-3}	2.3225	2.3403
6×10^{-7}	1.1854	1.1931	4×10^{-3}	2.6526	2.6713
1×10^{-6}	1.3074	1.2936	6×10^{-3}	2.8672	2.8816
2×10^{-6}	1.4933	1.4760	1×10^{-2}	3.1623	3.1611
4×10^{-6}	1.7057	1.7019	2×10^{-2}	3.6212	3.6776
6×10^{-6}	1.8436	1.8481	4×10^{-2}	4.1255	4.0368
1×10^{-5}	2.0333	2.0453			

The approximately linear portion invariably forms part of the logarithmic curve calculated from the multi-molecular layer adsorption formula, whatever values of the exponential terms are taken, provided that the exponential term of the first layer is much greater than that of the second layer, and the exponential terms of the second, third, etc. layers progressively decrease. The multi-molecular layer adsorption formula, therefore, offers an explanation of the Freundlich isotherm and indicates that this isotherm is a specific characteristic of such adsorption.

It has now been shown that the equilibrium for each layer of the multi-layer adsorption formula is identical in form with the Langmuir expression and that the equilibria give the Freundlich isotherm. These facts, however, do not necessarily establish full confidence in the formula. If a clean surface be exposed to a gas at a temperature below the critical temperature of that gas and if the pressure of the gas be steadily increased, the adsorption will increase until the gas liquefies. The supreme test of any adsorption formula, therefore, is that it must pass without any break of continuity whatever into an equation for the liquid state.

Now the equilibrium equation for any layer of adsorbed gas can be written in the form

$$A = \frac{k_2}{k_1} \cdot \frac{t}{s-t} e^{-Q/RT}, \quad (7)$$

where t is the fraction of the underlying layer s (where s is itself a fraction of the adsorbed layer below) which is coated by the next outer layer. It may be assumed that this equation defines the concentration of the outermost adsorbed layer when the molecular concentration of the gas at equilibrium is A . On applying this equation to any non-associating liquid, such as CO_2 , the vapour pressure of the liquid, p , is expressed by the equation

$$\frac{p}{p_c} = \frac{t}{1-t} e^{-Q/RT}, \quad (8)$$

where p_c is the critical pressure of the gas and Q is the molecular latent heat of the liquid at the temperature T . This equation is of considerable interest. It indicates that in the adsorption formulae the molecular concentration of the gas A , which is a fraction, is the ratio of the equilibrium pressure to the critical pressure of the gas. Furthermore, the equation indicates that in the case of liquids the term k_2/k_1 of the adsorption formula becomes equal to unity.

The most interesting feature of equation (8) is the term $t/(1-t)$ which has replaced the term $t/(s-t)$ in the adsorption equation. The term $t/(1-t)$

indicates that in a liquid the layer of molecules immediately below the outermost layer is complete. In this way the surface of a liquid is physically defined as a complete layer of molecules, a fraction of which is coated with an outer layer of molecules. This fraction t is a function of the temperature and decreases with rise in temperature. Equation (8), therefore, may be accepted as the criterion of the liquid state.

It is generally accepted that the latent heat of a liquid becomes zero at the critical temperature. This, however, cannot be true if equation (8) is correct.

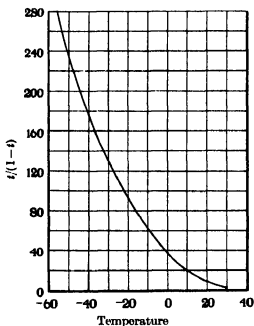


FIG. 2

If any liquid exists at the critical point, however small in amount, it must require energy for its conversion into gas. If no liquid exists at the critical point, it is absurd to speak of latent heat.

As is shown in fig. 2, the values of $t/(1-t)$ for liquid CO_2 , when plotted against the temperature lie on a perfectly smooth curve and this curve when continued to the critical point, 31°C ., gives $t/(1-t) = 2.00$. It follows from this that at the critical point the liquid state must exist and that equation (8) of the liquid state becomes at the critical point

$$\frac{p}{p_c} = 1 = 2 \times e^{-2}, \quad (9)$$

where $x = 0.6931478$ and is a fundamental constant which is independent of the nature of the substance. It may be noted that 0.6931478 is the value of Q_c/RT_c , where Q_c is the molecular latent heat required to separate the uni-molecular layers and T_c is the critical temperature. The critical temperature of any substance, therefore, may be defined as the temperature T at which $Q_c/RT_c = 0.6931478$.

Since it is evident from equation (8) that the term $t/(1-t)$ is an exponential, the correct form of equation (9) for the liquid state at the critical temperature is

$$\frac{p}{p_c} = e^{0.6931478} \times e^{-0.6931478} \quad (10)$$

In considering the modification of this equation so that it may give the values of p/p_c over the whole range of temperature from the critical point to the triple point, it may be accepted without argument that the energy required to remove molecules from within the body of a liquid at a given temperature is a function of the relative volume of the liquid at that temperature. If the relative volume V of a liquid be defined as v_r/v_l , where v_r is the specific volume of the liquid at the critical temperature and v_l is the specific volume at the temperature of observation, the molecular latent heat of a liquid will be $0.6931478 \times \phi_2(V)$. Similarly, the exponent of the first exponential will be $0.6931478 \times \phi_1(V)$. The general equation for the liquid state thus becomes

$$\frac{p}{p_c} = e^{0.6931478 \times \phi_1(V)} \times e^{-0.6931478 \times \phi_2(V)} \quad (11)$$

The attraction between two molecules within the body of a liquid, or, conversely, the energy required to separate them varies inversely with a function of the distance between them, this distance being proportional to the cube root of the volume. It follows from this that $\phi_2(V)$ in the second exponential may be expressed by $V^{a/3}$, where a depends on the properties of the substance. It is not possible at this stage to define the value of $\phi_1(V)$ in the first exponential of equation (11), but it may be assumed that $\phi_1(V)$ is V^b , where b is smaller than $a/3$. The final form of the equation of the liquid state, therefore, is

$$\frac{p}{p_c} = e^{0.6931478V^b} \times e^{-0.6931478V^{a/3}} \quad (12)$$

In order to test this formula, strictly accurate values of the three variables, vapour pressure, specific volume, and latent heat of a non-associating liquid from the triple point up to 1° below the critical point and of the specific volume at the critical point are required. The only case known to me

is CO_2 , the specific volumes and latent heats of which have been determined up to 1° below the critical point. Considerable encouragement was experienced at the outset when it was found that the latent heats of liquid CO_2 from -56.6 to 20°C. are expressed with reasonable approximation by $0.6931478V^{8/3}$.

The agreement between the calculated and observed values is sufficiently close to justify the provisional acceptance of $0.6931478V^{8/3}$ as the value of the exponent in the second exponential of equation (12) when applied to liquid CO_2 .

When the value of b in the equation was calculated from

$$e^{0.6931478V^b} = \frac{p}{p_c} \times e^{0.6931478V^{8/3}},$$

it was found that b is not constant. The value of b increases from 2.2547 at -56.6°C. to 2.6004 at 30°C. , and it is of interest to note that close to the critical point the value of b has very nearly reached $8/3$. This is in accordance with expectation, since equation (12) indicates that close to the critical point when p/p_c is only a little less than unity, b should be very nearly equal to $a/3$, that is $8/3$ in the case of CO_2 .

There can be no doubt that the increase in the value of b with rise in temperature is due to the fact that the surface/volume ratio increases with decrease in the relative volume V . In spite of searching endeavour it was not found possible to evaluate the surface/volume ratio from the observed values of the specific volumes v_l and v_c . These values are in fact open to criticism, since the observed values of v_l for liquid CO_2 do not lie on a smooth curve, and it is not necessary to emphasize the fact that small errors in the value of V are greatly magnified in the exponential formula. It is of some interest, however, to note the fact that the progressive change in the surface/volume ratio can be compensated to a surprising extent by increasing the value of v_c , the specific volume at the critical point, in the first exponential of equation (12). It was found that the following formula holds with a reasonable degree of approximation for CO_2 between -56.6° and -15°C. ,

$$\frac{p}{p_c} = e^{0.6931478V_1^{1/3}} \times e^{-0.6931478V_1^{8/3}}, \quad (13)$$

where $V_1 = 2.67763/v_l$ and V_2 has the same values as previously, namely, $2.1547/v_l$.

Now the evidence that has been obtained indicates very definitely that there must be some inherent error in the observed values of the specific volumes. If this is correct, the probability of error increases at the higher

temperatures and is greatest in the case of the critical volume v_c . On the assumption that the observed value of v_c , 2.1547 is too small, it was at once found that the latent heats of liquid CO_2 between -56.6° and -5°C. are more accurately given by $0.6931478V^{7.703/3}$ when the value of v_c is increased from 2.1547 to 2.228411. The values of the latent heat calculated from this expression and from the previous formula are given in the third and fourth columns of Table II and there can be no question as to the superiority of the new formula.

TABLE II—LATENT HEAT OF LIQUID CO_2

Temp. ° C.	Calc.			Temp. ° C.	Calc.		
	Obs.	New	Old		Obs.	New	Old
-56	81.1	80.96	81.4	-25	69.3	69.47	68.9
-55	80.6	80.51	80.9	-20	67.0	67.09	66.6
-50	78.9	79.15	79.5	-15	64.3	64.23	63.6
-45	77.2	77.54	77.7	-10	61.4	61.51	60.8
-40	75.5	75.70	75.2	-5	58.3	58.22	57.0
-35	73.6	73.83	73.7	0	55.0	54.79	53.8
-30	71.5	71.74	71.5	5	46.4	51.11	50.0

It may be noted that above -5°C. the calculated values of the latent heat become increasingly larger than the observed values. This is doubtless due to the fact that the error in the critical volume is due to increasing errors in the observed values of v_l from 0 to 30°C. This probability need not be discussed, since the only assumption made is that v_l is 2.228411 instead of 2.1547. Now the improvement found in the new exponent of the second exponential, $0.6931478V^{2.5676}$ enables us for the first time to consider the surface/volume ratio. The possibility suggested itself that at low temperatures, when the relative volume V is large, a liquid may have the properties of a sphere, the surface/volume ratio of which is inversely proportional to the cube root of the volume. It was at once found that this is correct and that the vapour pressures and latent heats of liquid CO_2 from -56.6 to -26°C. are accurately expressed by the equation

$$\frac{p}{p_c} = e^{0.6931478V^{2.5676}} \times e^{-0.6931478V^{2.5676}}, \quad (14)$$

where $a = 1.165806 - \sqrt[3]{V}$ and $V = 2.228411/v_l$.

In view of the magnification of small errors in the observed values of v_l it was found preferable in applying this equation to calculate the values of v_l which give the observed values of the vapour pressure. The results are given in Table III.

TABLE III

Temp. ° C.	Specific volume		Vapour pressure in mm.		Latent heat in cal./g.	
	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.
-56.6	0.8474049	0.8482	3885.4	3885.2	81.16	81.1
-56	0.8491122	0.8496	3987.9	3987.9	80.96	—
-55	0.8519630	0.8525	4163.2	4163.2	80.64	80.6
-54	0.8548313	0.8554	4344.3	4344.3	80.31	—
-53	0.8577128	0.8576	4531.1	4531.1	79.99	—
-52	0.8606000	0.8606	4723.0	4723.0	79.66	—
-51	0.8635267	0.8636	4922.7	4922.7	79.33	—
-50	0.8664620	0.8658	5127.8	5127.8	78.99	78.9
-49	0.8694100	0.8688	5338.5	5339	78.66	—
-48	0.8723861	0.8718	5557.3	5557	78.32	—
-47	0.8753720	0.8749	5781.0	5781	77.98	—
-46	0.8783813	0.8780	6011.9	6012	77.64	—
-45	0.8814171	0.8803	6250.1	6250	77.29	77.2
-44	0.8844600	0.8834	6494.0	6494	76.95	—
-43	0.8875400	0.8865	6746.1	6746	76.59	—
-42	0.8906342	0.8897	7004.6	7005	76.24	—
-41	0.8937557	0.8929	7271.0	7271	75.89	—
-40	0.8969050	0.8961	7545.0	7545	75.53	75.5
-39	0.9000735	0.8993	7825.9	7826	75.17	—
-38	0.9032724	0.9025	8115.0	8115	74.81	—
-37	0.9065038	0.9058	8412.3	8412	74.44	—
-36	0.9097438	0.9091	8716.0	8716	74.07	—
-35	0.9130270	0.9124	9029.0	9029	73.72	73.6
-34	0.9163374	0.9158	9349.9	9350	73.33	—
-33	0.9196759	0.9191	9679.0	9679	72.95	—
-32	0.9230505	0.9226	10017.0	10017	72.57	—
-31	0.9264523	0.9268	10363.0	10363	72.17	—
-30	0.9299000	0.9302	10718.0	10718	71.80	71.5
-29	0.9333638	0.9337	11082.0	11082	71.40	—
-28	0.9368746	0.9381	11455.0	11455	71.01	—
-27	0.9404297	0.9416	11838.0	11838	70.61	—
-26	0.9440121	0.9452	12229.0	12229	70.21	—
-25	0.9476395	0.9497	12630.0	12630	69.80	69.3

It may be noted in the first place that the latent heats given in the table were calculated from the calculated values of v_1 given in the second column. The agreement between these and the observed values is remarkably good. The real test of the validity of equation (14) is the comparison of the calculated and observed values of the specific volume. In view of the importance of this test, all the data are collected together in Table IV.

The last column of that table contains the differences between the consecutive values of the observed specific volumes for each degree rise in temperature, and these differences fully justify the statement previously

made that the observed values do not lie on a smooth curve. A comparison between the last two columns of Table IV shows that equation (14) passes the test very satisfactorily. In every case where the difference between the calculated and observed values of the specific volume undergoes a marked change, such change is associated with a discrepancy in the last column of the table. It may, therefore, be claimed that equation (14) has been proved to apply to liquid CO_2 between -56.6 and -26°C. , that is to say over about one-third of the temperature range over which liquid CO_2 exists

TABLE IV

Temp. $^\circ \text{C.}$	v_1 calc.	v_1 obs.	v_1 obs. - v_1 calc.	Δv_1 obs.
-56.6	0.8474049	0.8482	+ 0.0007851	0.0014
-56	0.8491122	0.8496	+ 0.0004878	0.0029
-55	0.8519630	0.8525	+ 0.0005370	0.0029
-54	0.8548313	0.8554	+ 0.0005687	0.0022
-53	0.8577128	0.8576	- 0.0001128	0.0030
-52	0.8606000	0.8606	\pm 0.0000000	0.0030
-51	0.8635267	0.8636	+ 0.0000733	0.0022
-50	0.8664620	0.8658	- 0.0006620	0.0030
-49	0.8694100	0.8688	- 0.0006100	0.0030
-48	0.8723861	0.8718	- 0.0005861	0.0031
-47	0.8753720	0.8749	- 0.0004720	0.0031
-46	0.8783813	0.8780	- 0.0003813	0.0023
-45	0.8814171	0.8803	- 0.0011171	0.0031
-44	0.8844800	0.8834	- 0.0010600	0.0031
-43	0.8875400	0.8865	- 0.0010400	0.0032
-42	0.8906342	0.8897	- 0.0009342	0.0032
-41	0.8937557	0.8929	- 0.0008557	0.0032
-40	0.8969050	0.8961	- 0.0008050	0.0032
-39	0.9000735	0.8993	- 0.0007735	0.0032
-38	0.9032724	0.9025	- 0.0007724	0.0033
-37	0.9065038	0.9058	- 0.0007038	0.0033
-36	0.9097438	0.9091	- 0.0006438	0.0033
-35	0.9130270	0.9124	- 0.0006270	0.0034
-34	0.9163374	0.9158	- 0.0005374	0.0033
-33	0.9196759	0.9191	- 0.0005739	0.0035
-32	0.9230505	0.9226	- 0.0004505	0.0042
-31	0.9264523	0.9258	+ 0.0003477	0.0042
-30	0.9299000	0.9302	+ 0.0003000	0.0035
-29	0.9333838	0.9337	+ 0.0003362	0.0044
-28	0.9368746	0.9381	+ 0.0012254	0.0035
-27	0.9404297	0.9416	+ 0.0011703	0.0036
-26	0.9440121	0.9452	+ 0.0011879	0.0045
-25	0.9476395	0.9497	+ 0.0020605	

Above 24°C. equation (14) commences to fail rapidly and it is probable that this is due to the surface/volume ratio changing to a different function of the relative volume at the higher temperatures. There is, however, no

need to deal with this problem in the present communication. As was stated previously, the supreme test of a gaseous adsorption formula is that below the critical temperature of a gas the formula must, when the pressure is increased sufficiently, change without any break of continuity into an equation of the liquid state. There can be little doubt that the multilayer adsorption formula has passed this test satisfactorily and hence it may be discussed with some confidence in its correctness.

It is to be noted in the first place that the term A in the adsorption formula is to be replaced by p/p_c , where p is the equilibrium pressure and p_c is the critical pressure of the adsorbed gas.

In the second place it is evident that the Freundlich isotherm has no physical significance in connexion with adsorption and is a mathematical property of the sums to n terms of the type of diminishing geometrical series which govern multilayer adsorption. For this reason the slope of the isotherm is of value in that it is determined by three factors which in order of importance are the ratio of the exponential terms of the first and second layers, the number of the layers, and the ratios of the exponential terms of the second, third, fourth, etc., layers. The effect of decreasing the ratio of the exponential terms of the first and second terms in an eight-layer formula is shown in Table V.

TABLE V

First exponential	Ratio of first to second exponential	Slope of isotherm
2×10^4	1000	0.37273
4×10^4	200	0.3740
2×10^4	100	0.3750
1×10^4	50	0.843
4×10^3	20	0.932
2×10^3	10	0.9415

It is to be noted that the pressure range over which the isotherm holds is progressively decreased as the ratio of the first and second terms becomes smaller. When this ratio is 50 or less the amount of gas adsorbed is at first proportional to the pressure and this region is followed by a short pressure range over which the isotherm relation holds. This is of some interest, since Freundlich pointed out that these two relations are given by argon (Freundlich 1930, p. 157).

The isotherm should, therefore, prove of service in evaluating the exponential terms of an adsorbed system. Additional information may be obtained from the value of Q_c of the adsorbed gas, since the value of Q/RT for the outermost adsorbed layer must be greater than Q_c/RT .

Heats of adsorption and de-sorption may now be considered. Since the formation of the adsorption complexes is *ex hypothesi* an isothermal process, the heat of adsorption, except in cases where chemical change occurs, should in general be the same for all gases, namely that due to the loss of free energy by the gas.

The heat of desorption is the energy Q which is necessary to dissociate the adsorption complexes and is usually determined by the differential method. In discussing this method and its inherent errors the case of the adsorption of a unimolecular layer of gas on unit area of surface may be considered first. Equation (1) can in this case be written in the form

$$\frac{p_1}{p_c} = \frac{t}{1-t} \times e^{-Q/RT_1},$$

where p_1 is the equilibrium pressure, p_c is the critical pressure of the gas and t is the fraction of the unit surface area coated by the gas, it being assumed that as in the equation of the liquid state $k_a/k_1 = 1$. Let the temperature now be raised from T_1 to T_2 , then if $\frac{t}{1-t}$ remain constant

$$\frac{p_2}{p_c} = \frac{t}{1-t} \times e^{-Q/RT_2}$$

and

$$\ln \frac{p_1}{p_2} = \frac{Q}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right).$$

This, however, is an ideal condition which cannot be realised experimentally for the reason that it would necessitate that the whole apparatus including the pressure gauge have the same volume as that of the adsorbent. It is obvious that in practice the observed increase in pressure would be smaller than the true value, because any rise in temperature must result in a decrease in the value of $t/(1-t)$. This decrease is determined by the ratio of the total volume of the apparatus, less the volume of the adsorbent, to the volume of the gas adsorbed on the adsorbent at the lower temperature.

The simplest method of determining the error caused by the assumption that $t/(1-t)$ remains constant when the temperature is raised is a graphical one. If the adsorption equilibrium be expressed by

$$\log \frac{t}{1-t} = \log \frac{p}{p_c} + \frac{Q'}{T},$$

where

$$Q' = Q/2.303R,$$

then for each value of T the relation between $\log t/(1-t)$ and $\log p/p_c$ will be

expressed by a straight line at an angle of 45° to the $\log t/(1-t)$ and $\log p/p_e$ axes. On the assumption that the two temperatures at which pressure measurements are to be made are $T_1 = 298^\circ \text{ C.}$ and $T_2 = 308^\circ \text{ C.}$, the isotherms for these two temperatures are shown in fig. 3, where

$$AE = Q'(\frac{1}{298} - \frac{1}{308}).$$

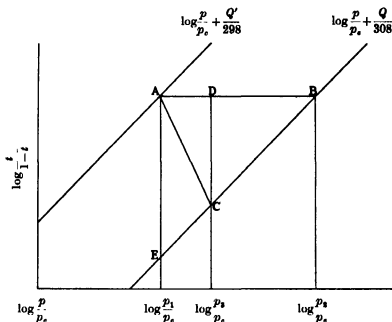


FIG. 3

The equilibrium state at 298° C. is expressed by

$$\log \frac{t}{1-t} = \log \frac{p_1}{p_e} + \frac{Q'}{298},$$

and this equilibrium may be denoted by the point *A* on the 298° C. isotherm. Let it first be assumed that the total volume of the apparatus is equal to the volume of the adsorbent, then if the temperature be raised to 308° C. the new equilibrium will be the point *B* on the 308° C. isotherm, this equilibrium being

$$\log \frac{t}{1-t} = \log \frac{p_2}{p_e} + \frac{Q'}{308}.$$

Now let the volume be increased isothermally at 308° C. to *V*, where *V* is the total volume of the apparatus, less the volume of the adsorbent. The new

equilibrium must lie on the 308°C. isotherm and will be established by an equal decrease in $\log t/(1-t)$ and $\log p/p_e$, it being manifest that no increase in pressure could take place unless some of the adsorbed gas were desorbed. The new equilibrium may be represented by the point C and the problem is to find the position of C . The position of C is geometrically determined by the ratio CD/AD or the tangent of the angle CAD . Experimentally the position of C is determined by the ratio V/v , where v is the volume of the gas adsorbed at the lower temperature. It is reasonable, therefore, to assume that CD is proportional to V and that AD is proportional to v , that is to say, $CD = xV$, $AD = xv$ and $\frac{CD}{AD} = \frac{V}{v}$. Since $BD = CD$, it follows that

$$AD = AB \times \frac{AD}{CD + AD},$$

and

$$AD = AB \times \frac{v}{V + v}.$$

Now $AB = \log p_2 - \log p_1$, where p_2 is the pressure which would have been observed at 308°C. under the ideal condition that $V = 0$, and $AD = \log p_3 - \log p_1$, where p_3 is the pressure which is observed experimentally. The conclusion is reached, therefore, that the value of Q determined experimentally is the true value of Q multiplied by $v/(V + v)$.

In order to gain some idea of the magnitude of the term $v/(V + v)$ the measurements made in these laboratories with purified "Superfloss" kieselguhr may be instanced. Since these measurements will shortly be communicated in detail, it will suffice to say that 100 g. of this kieselguhr are completely coated by 0.00544 g. mol. of NiO as a unimolecular layer, the surface area being 2.869×10^6 sq. cm. For the present purpose it may be assumed that a unimolecular layer of an adsorbed gas is the same, i.e. that $v = 0.00544 \times 22,412 = 121.92$ c.c. The density of the kieselguhr is 2.3 and the volume of 100 g. is 43.5 c.c. If the total volume of the apparatus is 250 c.c., $V = 206.5$ and

$$\frac{v}{V + v} = \frac{121.92}{328.42} = 0.37123,$$

that is to say the volume of Q found experimentally will be a little greater than one-third of the true value.

Since the experimental error varies directly with V and inversely with v , it may become very large indeed. It is also to be noted that, since v decreases with the equilibrium pressure, the value of Q determined experimentally will decrease with the pressure.

It is a simple matter to find by the graphical method the true value of Q from the observed values of p_1 and p_2 at two different temperatures, since both V and v can be found experimentally, and the method is shown in fig. 4. Two parallel straight lines are drawn at an angle of 45° to the $\log t/(1-t)$ and $\log p/p_c$ axes, so that the length AB on the straight line AC parallel to the $\log p/p_c$ axis is equal to $\log p_2 - \log p_1$. On the perpendicular dropped from B to the $\log p/p_c$ axis the point D is fixed so that $BD/AB = V/v$ or, alternatively, the angle DAB is drawn, the tangent of which is V/v . A straight line is then drawn through D at an angle of 45° to the $\log p/p_c$ axis

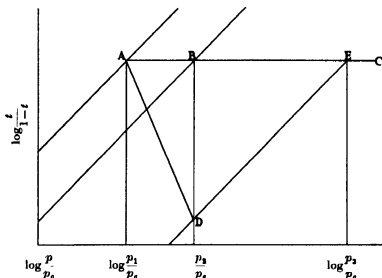


FIG. 4

and this line cuts the straight line AC at E . It follows from the previous argument that $AE = \log p_2 - \log p_1$, where p_2 is the pressure which would have been observed under the ideal condition that $V = 0$. Since $BE = BD$,

$$AE = AB \times \frac{AB + BD}{AB},$$

and hence

$$\log \frac{p_2}{p_1} = \log \frac{p_2}{p_1} \times \frac{V+v}{v}.$$

The value of Q calculated on the usual assumption that the amount of gas adsorbed remains constant at the higher temperature must therefore be multiplied by $(V+v)/v$. It is important to note that this factor is independent of the temperature interval $T_2 - T_1$.

The normal condition under which several layers of gas are adsorbed may now be considered. In this case the value of Q for the whole system is of great interest, because each separate layer is characterised by its own exponential term. It is a simple matter to calculate the true value of Q if the exponential terms of all the adsorbed layers are known. In order to illustrate the method, an eight-layer system may be taken having the following series of exponential terms: -2×10^5 , 200, 100, 60, 40, 30, 24 and 20; and it may be assumed that these are the values at 25°C . The values of Q for the eight layers are then 7233, 3140, 2729, 2426, 2186, 2015, 1883, and 1775 cal., respectively. It may also be assumed that the higher temperature selected is 35°C ., and at that temperature the values of the exponentials are 1.3456×10^5 , 168.39, 86.112, 52.531, 35.485, 26.863, 21.647 and 18.268. The values of α , or the total amount of gas adsorbed on unit area of the surface, are then calculated for a number of values of p/p_e at 25 and 35°C . The values of $\log \alpha$ are then plotted against $\log p/p_e$, and the two $\log \alpha$, $\log p/p_e$ curves for 25 and 35°C . drawn. It will then be found that the values of $\log p_2 - \log p_1$ for equal values of $\log \alpha$ on the two curves progressively increase with decrease in p/p_e up to a maximum value at $\log p/p_e = 8.000$. This maximum value of $\log p_2 - \log p_1$ is 0.1718 and gives $Q = 7221$ cal. which is very nearly equal to the value of Q for the first adsorbed layer. This is readily explained by the fact that when the equilibrium pressure is very small the adsorbed gas is almost entirely in the first layer.

The fall in the value of Q of the whole system as the equilibrium pressure is increased is due to the fact that the amount of gas adsorbed in the second, third, etc., layers is progressively increased, with the result that the decreasing values of Q for each of these layers causes the value of Q for the whole system to be progressively decreased. It is to be noted that as the equilibrium pressure is increased the fall in the value of Q for the whole system is continuous and shows no evidence of any periodicity between the two extremes of $\log p/p_e = 8.000$ and $\log p/p_e = 0.0$. At the critical pressure, when $p/p_e = 1$, $\log p_2 - \log p_1 = 0.0690$ which gives 2900 cal. as the value of Q for the whole system. This value is about midway between the values of Q characteristic of the second and third layers and it follows from this that the values of Q for the outer layers cannot be determined by the differential method.

There is little doubt that all experimental determinations of Q for the whole system in the case of multilayer adsorption must also be multiplied by the factor $(V + v)/v$. In view of the fact that in the case of a unimolecular adsorbed layer the experimental value of Q decreases with decrease in the equilibrium pressure, the experimental evidence that Q increases with

decrease in pressure affords a singularly strong argument in favour of the multilayer adsorption formula.

The exponential term of the first adsorbed layer may now be discussed in some detail. The magnitude of this term is of fundamental importance since it governs the phenomena of preferential adsorption, poisoning of surfaces and chemi-sorption. As was stated in the development of the multilayer formula, the energy term Q_1 of the first adsorbed layer is equal to one or more rotation or rotation-vibration quanta which are donated by each adsorbing unit in the surface layer of the adsorbent to each molecule of the adsorbate. Q_1 , therefore, may be as large as 80,000–100,000 cal. or as small as a few hundred calories. In special cases, such as that of helium, Q_1 is relatively very small. It at once becomes obvious that the phenomena observed in adsorption will be of very varied types, embracing such extremes as chemi-sorption and the so-called van der Waals adsorption.

In comparing the stability of systems in which a single layer only is adsorbed it is convenient to adopt as the standard of comparison the equilibrium pressure when half the surface area of the adsorbent is coated,

that is to say when $t/(1-t) = 1$. Under these conditions $\frac{1}{p} = \frac{e^{Q_1/RT}}{p_c}$, and

hence the stability is proportional to the exponential term divided by the critical pressure. It follows from this that preferential adsorption is governed entirely by Q_1 , since out of a mixture of gases that gas will be adsorbed for which the value of Q_1 is the greatest.

This is of some importance, since the surface of an adsorbent which is to be used in experimental work must be coated by the substance having the largest value of Q_1 with which it has come into contact. If Q_1 for this adsorbed substance be 40,000 cal. and p_c be 70 atmospheres, the equilibrium pressure when half the surface is coated will be 2.6×10^{-25} at 25° and 2.65×10^{-7} at 500° . The complete removal of all adsorbed impurities from a surface, therefore, will necessitate very prolonged heating and the use of an efficient molecular exhaust pump.

The most important significance of the value of Q_1 for the first layer is to be found in the phenomenon of chemi-sorption. Since the word chemi-sorption means the association of some chemical change with the act of adsorption, the correct definition of chemi-sorption is that adsorption in which Q_1 of the first adsorbed layer is equal to or greater than either the energy of activation or the heat of formation of the adsorbate. It must be noted that this definition by no means necessitates that Q_1 should be very large, since in some cases Q_1 may be small. There are in fact four possibilities and these may be illustrated by examples.

The most familiar case in which Q_1 is equal to or greater than the energy of activation is that of hydrogen peroxide which is decomposed when adsorbed on any surface. Apart from any other evidence, it is evident that in this case Q_1 is small, since it is a simple matter temporarily to poison the surface by the use of a so-called preservative in small concentration.

The second possibility is that Q_1 is equal to or greater than the heat of formation of the adsorbate. This may be illustrated by the adsorption of hydrogen sulphide, the heat of formation of which is about 3000 cal. When Q_1 is equal to or greater than 3000 cal., the first layer of hydrogen sulphide adsorbed will be decomposed into free hydrogen and a unimolecular layer of sulphur which remains adsorbed on the surface. The result of this will be that the surface becomes completely poisoned, since an adsorbed layer of sulphur is exceedingly difficult to remove.

The third possibility is when the surface molecules or atoms undergo a chemical reaction either with a product of decomposition of the adsorbate or with the activated adsorbate molecules. The first alternative may be illustrated by the adsorption of hydrogen sulphide on a metal surface when as the result of Q_1 being large enough to decompose the first layer adsorbed a unimolecular layer of the metallic sulphide is formed. This is the most serious type of poisoning that can occur. The second alternative in which the activated adsorbate reacts with the atoms or molecules in the surface may be illustrated by the adsorption of oxygen by charcoal. In this case the adsorption complex $C-O_2^+$ undergoes a change into the molecule CO_2 which remains adsorbed on the second layer of C atoms. The fact that CO_2 is always found when the adsorbed gases are removed at room temperature is a proof that the type of chemical change specified has taken place. A similar example of this is afforded by the formation of a unimolecular layer of metallic hydride when hydrogen is adsorbed on certain metals.

It is obvious that when this type of chemi-sorption takes place the adsorption of the first layer will no longer be an isothermal process, since an abnormal evolution of energy will be observed. Indeed the evolution of an abnormal amount of energy during the adsorption of a gas may be accepted as evidence of the specified type of chemi-sorption.

The fourth possibility arises from a development of the theory of adsorption which for the sake of simplicity was not discussed when the formation of the adsorption complex was originally defined. No reference was made to the effect of temperature on the number of rotation or rotation-vibration quanta donated by the adsorbing unit in the surface to each molecule of the adsorbate. Since the rotational energy increases with rise of temperature, it is possible that a molecule of adsorbate in forming its

adsorption complex with an adsorbing unit of the adsorbent may accept a larger amount of energy at a high temperature than it does at a low temperature. Similarly, if a complex formed at a low temperature by the donation of Q_1 calories is raised to a higher temperature, it is possible that Q_1 may be increased by a definite amount. Such increase in Q_1 will not be a gradual one as the temperature is raised, but will take place suddenly when the temperature interval is increased to a definite extent. If such sudden increase from Q_1 to Q_2 take place in the case of the first adsorbed layer when the temperature is raised, the energy of activation Q_2 of the adsorbate may endow it with a reactivity which it did not possess at the lower temperature. In general, therefore, it would seem possible that with a given surface and gas the phenomena observed at two different temperatures may be essentially different.

As an example of this change the adsorption of CO_2 by charcoal may be considered. At room temperature the energy of activation of the adsorbed molecules is insufficient to fall within the definition of chemisorption. When the temperature is raised above a certain value it is known that CO is formed, and hence it is evident that the energy of activation has become sufficiently large for the chemical change



to take place.

The complete sequence of phenomena to be expected when oxygen is adsorbed by charcoal may now be described. When adsorption takes place at room temperature, the energy donated by the adsorbing carbon units to the oxygen molecules is sufficient to activate the latter with the result that a unimolecular layer of CO_2 is formed which remains adsorbed, and consequently during the adsorption of the first unimolecular layer of oxygen an abnormally large heat evolution will be observed. If the whole of the gas be now desorbed at room temperature, it will be found to be a mixture of CO_2 and O_2 and the volume of the CO_2 will give an approximate measure of the surface area of the charcoal, whilst the volume of O_2 divided by the volume of CO_2 will give an approximate measure of the number of layers of O_2 adsorbed on the first unimolecular layer of CO_2 .

If, before desorption is carried out, the temperature be raised above a critical limit, the increased energy of activation of the CO_2 in the first adsorbed layer will be sufficient for the change of the C-CO_2^+ adsorption complex into 2CO to take place and if the gas be now desorbed at the higher temperature it will be found to consist of O_2 , CO and CO_2 , the volume of CO being greater than that of the CO_2 . The presence of any CO_2 might seem to be

unexpected, but it must be remembered that any adsorbed system is an equilibrium condition and that some of the adsorbed molecules of CO_2 in the first unimolecular layer will be replaced by O_2 molecules, with the result that some CO_2 will always be found in the gas phase.

The sequence of phenomena actually observed is sufficiently similar to that described above to justify full confidence in the extension of the theory of adsorption.

CONCLUSIONS

1—The adsorption of a gas on a surface is due to the formation of adsorption complexes $S-A^+$ in which an adsorbing unit S in the surface has donated one or more rotation or rotation-vibration quanta to a molecule of the adsorbate.

2—The adsorption complex is dissociated by the supply of an amount of energy which is equal to that donated by the adsorbing unit.

3—The fraction r of the total surface area coated by the first unimolecular layer is expressed by

$$r = \frac{\frac{p}{p_c} e^{Q_1/RT}}{1 + \frac{p}{p_c} e^{Q_1/RT}},$$

where p is the equilibrium pressure, p_c is the critical pressure of the gas and Q_1 is the donated energy. This expression is identical in form with the Langmuir formula.

4—The activated molecules of gas in the first adsorbed layer can donate energy to form a second adsorbed layer of gas and the fraction s of the first layer thus coated is expressed by

$$s = r \times \frac{\frac{p}{p_c} e^{Q_2/RT}}{1 + \frac{p}{p_c} e^{Q_2/RT}},$$

where Q_2 is the donated energy and is smaller than Q_1 .

5—Each successive adsorbed layer is expressed by a similar formula and the total amount of gas adsorbed is the sum of that in each layer,

$$r + s + t + \text{etc.}$$

6—Over a limited range of pressure the relation between the total amount of gas adsorbed and the equilibrium pressure is expressed by the Freundlich isotherm.

7—The Freundlich isotherm has no physical significance in adsorption and is a mathematical property of the sums to n terms of the particular type of diminishing geometrical series governing multilayer adsorption.

8—The supreme test of an adsorption formula is that below the critical temperature it must, when the pressure is sufficiently increased, pass without any discontinuity into an equation of the liquid state. The multilayer adsorption formula gives at the critical point

$$\frac{p}{p_c} = e^{0.6931478} \times e^{-0.6931478},$$

and from this the fundamental equation of the liquid state is found to be

$$\frac{p}{p_c} = e^{0.6931478V^b} \times e^{-0.6931478V^{a/3}},$$

where V is the relative volume of the liquid, a depends on the properties of the substance, b is the product of $a/3$ into a function of the surface-volume ratio of the liquid, and $0.6931478V^{a/3}$ is the molecular latent heat.

9—The vapour pressures and latent heats of liquid CO_2 between -56.6 and -26°C. are accurately expressed by the equation when $a/3 = 2.5676$ and $b = 2.5676 \times 1.165806 - \sqrt[3]{V}$.

10—Since the formation of the adsorption complexes is an isothermal process, the heat of adsorption should in general be the same for all gases, namely the free energy lost by the gas, provided that no chemical change takes place.

11—The heat of desorption is the energy needed to dissociate the adsorption complexes. When this energy is determined by the difference between the equilibrium pressures at two different temperatures the value found of Q is the true value of Q multiplied by the factor $v/(V+v)$, where v is the volume of gas adsorbed at the lower temperature and V is the total volume of the apparatus, less the volume of the adsorbent. Since v decreases with the pressure, the value of Q experimentally found for a unimolecular layer of adsorbed gas will also decrease with the pressure.

12—With multilayer adsorption the true value of Q for the whole system is a maximum at very low pressures and decreases as the equilibrium pressure is increased, owing to the fact that the values of Q for the second, third, etc., layers progressively decrease. The fact that the values of Q found experimentally increase with decrease in the pressure is a proof that the adsorbed state consists of several layers of gas.

13—The value of Q_1 for the first adsorbed layer governs the phenomena of preferential adsorption, poisoning and chemisorption.

14—Chemi-sorption is adsorption in which Q_1 for the first adsorbed layer is equal to or greater than either the energy of activation or the heat of formation of the adsorbate. When as the result of chemi-sorption a chemical reaction takes place between the adsorbing units in the surface and the molecules of the adsorbate, abnormal heats of adsorption are observed.

REFERENCES

- Baly, E. C. C. 1928 *Rep. Brit. Ass.* p. 35.
 — 1935 *Nature, Lond.*, 136, 28.
 — 1936 *J. Soc. Chem. Ind.* 55, 7.
 Freundlich, H. 1930 "Kapillarchemie," 1, 153-68, 4th ed. Akademische Verlagsgesellschaft, Leipzig.

The Band Systems ending on the $1s\sigma 2s\sigma^1\Sigma_g(1X_g)$ state of H_2 —Part I

BY O. W. RICHARDSON, F.R.S., *Yarrow Research Professor
of the Royal Society*

(Received 29 January 1937)

1—THE GENERAL PROBLEM

It is now well established that the electronic states of the band systems of H_2 have a close analogy to those of atomic helium and consist of a set of singlet states and a set of triplet states. There are no known combinations between singlet and triplet states. The ground level of H_2 is the $v = 0, K = 0$ level of the even state $1s\sigma 1s\sigma^1\Sigma_g$. The possible states with one electron excited to principal quantum number 2 are $1s\sigma 2s\sigma^1\Sigma_g$, $1s\sigma 2p\sigma^1\Sigma_u$, $1s\sigma 2p\pi^1\Pi_u$, $1s\sigma 2s\sigma^3\Sigma_g$, $1s\sigma 2p\sigma^3\Sigma_g$ and $1s\sigma 2p\pi^3\Pi_u$. Of these the only ones which can go down to the ground state are $1s\sigma 2p\sigma^1\Sigma_u$ and $1s\sigma 2p\pi^1\Pi_u$ on account of the triplet \longleftrightarrow singlet and odd \longleftrightarrow odd and even \longleftrightarrow even prohibitions. The bands with these transitions are well known and understood both in emission and absorption. A large number of emission band systems which go down to the states $1s\sigma 2p\sigma^1\Sigma_u$ and $1s\sigma 2p\pi^1\Pi_u$ from higher even states have been found and analysed so that we now have quite

an extensive knowledge of this part of the spectrum. We also have a comparable knowledge of higher states which go down to the triplet states $1s\sigma 2s\sigma^3\Sigma_g$ and $1s\sigma 2p\pi^3\Pi_u$. The odd triplet state $1s\sigma 2p\sigma^3\Sigma_u$ is believed on theoretical grounds to be unstable and non-existent. There should therefore be no bands with discrete lines which terminate on this state. Various facts bear out this contention.

This leaves only the even singlet $n = 2$ state $1s\sigma 2s\sigma^1\Sigma_g$ for further consideration. Until recently both the existence and identity of this state have been uncertain. For some time (Richardson 1934, p. 310) I have felt convinced that this state was identical with 1X , the lowest state which goes down to $1s\sigma 2p\sigma^1\Sigma_u$, in spite of some facts which appeared to point otherwise. In the last few months Dieke 1936 has found some important extensions of the $^1X \rightarrow 2p^1\Sigma$ bands in the extreme photographic infra-red. His analysis of these leaves no reasonable doubt that 1X is, in fact, $1s\sigma 2s\sigma^1\Sigma_g$. It was, I believe, he who first suggested this identification several years ago.

The ground state of 1X is so low (it lies some 25273 wave numbers below the ground state $1s\sigma^3\Sigma_g$ of the ion H_2^+) that there should be states with transitions to it giving rise to emission bands in the visible. None of the triplet states which go down to $2s^3\Sigma_g$ or to $2p^3\Pi_u$ can go down to 1X on account of the triplet \longleftrightarrow singlet prohibition. Neither can any of the numerous singlet states which go down to $2p^1\Sigma_u$ or to $2p^1\Pi_u$ on account of the even \longleftrightarrow even prohibition, as these are all even states like 1X . We should expect theoretically that any odd singlet state would be capable of transitions to or from 1X , in fact any state which is capable of transitions to or from the ground state of H_2 , which is also an even singlet state.

There are only three known states which undergo transitions to the ground state of H_2 . Two of these are $2p^1\Pi_u$ and $2p^1\Sigma_u$, already mentioned, and the other is a state called the D state by Hopfield. This state is the upper state of some bands found by him in the absorption spectrum of H_2 in the extreme ultra-violet beyond about 840 Å. As regards $2p^1\Pi_u$ this state lies some 25343 wave numbers below the ground state of H_2^+ and is therefore only just below 1X . Any transition between these two states would involve only a minute transfer of electronic energy, and any bands which might arise from such transitions would lie in the infra-red and in all probability well beyond the photographic infra-red limit. Richardson and Davidson's (1929a) determination of the depth of $2p^1\Sigma$ below H_2^+ , 34365 wave numbers, has recently been lowered by Beutler and Jünger (1936) to 34233. 1X thus lies some 8960 wave numbers above $2p^1\Sigma$, and the $^1X \rightarrow 2p^1\Sigma$ band system should lie in the photographic infra-red. These bands were discovered by Davidson and myself and independently by Finkelburg and Mecke in

1928, and have recently been extended to the band origin in the extreme photographic infra-red by Dieke (1936).

This brings us to the D state. What we can say about the bands which should arise from transitions from this state to 1X depends on the way in which we interpret Hopfield's observations. In 1932 I put forward the view (Richardson 1934, p. 303) that for some reason or another the $v' = 0, 1$ and 2 levels were missing from the progression as observed by Hopfield and that the bands should be renumbered starting with $v' = 3$ instead of $v' = 0$. This made the D state lie about 11560 wave numbers below $H_2 +$ with a Rydberg denominator 3.08, $\omega_0 = 2257$ and $x\omega_0 = 59.5$ (Richardson 1934, p. 325). I deduced from these numbers and other properties of the bands that this state was in all probability the $1s\sigma 3p\pi^1\Pi_u$ state, although it might with less likelihood be $1s\sigma 3p\sigma^1\Sigma_u$. Recently Hopfield's bands have been compared with the corresponding bands of D_2 by Beutler, Deubner and Jünger (1935). The comparison shows unquestionably that my vibrational numbering of the bands is correct and confirms the interpretation of this state as $1s\sigma 3p\pi^1\Pi_u$.

The expected location (band origin) of the $3p^1\Pi_u \rightarrow ^1X$ system could have been worked out on the basis of this interpretation of Hopfield's D state. Actually it was not, but the position of $3p^1\Pi_u$ was estimated by a comparison of the position of corresponding triplet and singlet states. This method is independent of the interpretation of the D state. After the bands have been described I shall show, however, that the properties of the upper state are consistent with its identification with the D state.

The actual method used to locate the $3p^1\Pi_u$ state was as follows: It is known that $2p^3\Pi_u$ lies some 291 wave numbers below $2s^3\Sigma_g$ and that $3p^3\Pi_u$ lies some 75 wave numbers below $3s^3\Sigma_g$. I now assume that the separation of the singlet states corresponding to these triplet states will be roughly in the same proportion. $2p^1\Pi_u$ lies some 70 wave numbers below $2s^1\Sigma_g$ (taken to be the same as 1X). Taking the 3^1O state to be $1s\sigma 3s\sigma^1\Sigma_g$ this lies some 11481 wave numbers below $H_2 +$. If the assumed proportionality occurs, this places $1s\sigma 2p\pi^1\Pi_u$ some 11500 wave numbers below $H_2 +$. Since 1X lies some 25272 below $H_2 +$ this puts the band origin in the neighbourhood of 13770 wave numbers. Having got the position of the band origin, we can construct the approximate position of the band lines from the known rotational and vibrational intervals of the final 1X level and the assumption that the upper state has properties with a reasonably close resemblance to the corresponding triplet state $3p^3\Pi_u$. The lines chosen have then to satisfy the known combination differences of the final 1X ($= 2s^1\Sigma$) state. For all these differences involving the $v = 0$ level of 1X , which he has

discovered, and most of those involving the $v = 1$ level, which he has greatly extended, I am indebted to a private communication from Professor Dieke.

Incidentally the success of this method of locating the position of $1s\sigma 3p\pi^1\Pi_u$ is to some extent also a confirmation of my identification of 3^1O as $1s\sigma 3s\sigma^1\Sigma_g$.

2—THE $1s\sigma 3p\pi^1\Pi_u \rightarrow 1s\sigma 2s\sigma^1\Sigma_g$ SYSTEM

The main strength of the band system lies in the axial sequence with the three principal bands $0 \rightarrow 0$, $1 \rightarrow 1$ and $2 \rightarrow 2$ all very close together and overlapping rather far down in the photographic infra-red around 7300 Å. It is probable that these axial bands are really rather strong, although Gale, Monk and Lee's eye estimates of the lines are not very high; since it is known from Davidson's work (cf. Richardson, Davidson, Marsden and Evans 1933, p. 69) that the lines in the infra-red with the same eye estimates have far more real energy than the lines in the visible. In fact, I think it highly probable that these bands contain a considerable proportion of all the lines of any strength in the whole H_2 spectrum which still remain unclassified. The true intensity of none of these lines has been measured.

The bands have P , Q and R branches. The wave numbers of the lines are set out in Table I followed by the eye intensity estimates of Gale, Monk and Lee (1928) (G), Poetker (1927) (P), Allibone (1926) (A), Merton and Barratt (1922) (MB), Tanaka (1925) (T) or Deodhar (1926) (D). All Poetker's wave numbers in the table and elsewhere throughout this paper have been reduced by 0.3 to take out the systematic difference between his measures and those of Gale, Monk and Lee. The final rotational combinations $R(K-1) - P(K+1)$ and the final vibrational combinations $P, Q, R(K) (v'' = v+1) - P, Q, R(K) (v'' = v)$ depend only on the energy levels of the final state 1X and have been got from the analysis of the initial state of the $^1X \rightarrow 2p^1\Sigma$ band system by Richardson and Davidson (1929*b*), Richardson (1934, 1935) and Dieke (1936). In Table I the combining $R(K-1)$ and $P(K+1)$ lines are listed in succession, and in the stronger bands the defects in the $R(K-1) - P(K+1)$ values from those got from the analysis of 1X in the $^1X \rightarrow 2p^1\Sigma$ system are shown on the left-hand side. In the weaker bands the defects, shown on the right, are those of the band lines from the wave numbers calculated from the corresponding lines of the stronger diagonal bands using the initial vibrational differences of $^1X \rightarrow 2p^1\Sigma$ which should be the same as the $P, Q, R(K) (v = v''+1) - P, Q, R(K) (v = v'')$ differences of the present bands. Where the combinations are between two single lines, both measured in

the table of Gale, Monk and Lee, the defects should not be much in excess of 0.10 wave number, but when the lines are taken from other tables the defects may be greater. There are also a considerable number of known and probably some unknown blends, and in such cases less accuracy is to be expected. However, there are not very many listed lines in the infra-red part of the spectrum where the stronger bands lie so that in most cases there is no doubt which are the band lines.

TABLE I

The 0 → 0 Band

^	R 0	13772.1(00) G, also 2 → 3, 3p ³ H → 2s ³ Σ, Q 4.
-0.26	P 2	13582.6(1) G, also Q 3 of the 1 → 1 band
^	R 1	13831.5(2a) G, also R 2 of the 2 → 2 band.
-0.05	P 3	13517.0(2) G.
^	R 2	13884.1(0a) G, also Q 5 of the 2 → 2 band.
-0.45	P 4	13447.8(2a) G, also 1 → 0, ¹ X → 2p ¹ Σ, R 1.
^	R 3	13931.0(5) G, also R 3 of the 2 → 2 band.
+0.45	P 5	13375.1(2) G, also 0 → 0, 3d ¹ A _g → 2p ¹ H, R 1.
^	R 3	13930.58(5) A.
+0.03	P 5	13375.1(2) G.
^	R 4	13971.0(2) G.
0	P 6	(13301.19) ab G.
^	R 5	14003.1(1) G, also 2 → 3, 3p ³ H → 2s ³ Σ, R 3.
-0.10	P 7	13224.6(0) G, also ? P 6 of the 1 → 1 band.
^	Q 1	13703.5(3) G.
	Q 2	13694.8(0b) G.
	Q 3	13681.9(1) G.
	Q 4	? 13663.3(0a) G, also R 0 of the 1 → 1 band.
	Q 5	13642.3(4) G, also P 1 of ? 0 → 0, x ¹ Σ _g → 2s ¹ Σ.
	Q 6	13611.6(0) G.

The 0 → 1 Band

R 0	(11444.14) ab P and G.
R 1	11505.77(0) P ab G, defect -0.12.
R 2	(11566.93) ab P and G.
R 3	11624.53(1) P ab G, defect -0.25.
R 4	11681.19(1b, h) P ab G, defect +0.06.
R 5	(11734.07) ? part of 11733.42(8) P (2) G, also 0 → 0, 3p ³ Σ → 2s ³ Σ, R 3.

The calculated positions of the P lines are P 2 = 11265.43, P 3 = 11210.28,

TABLE I—(continued)

$P 4 = 11158.05$, $P 5 = 11106.07$, $P 6 = 11069.30$ or 11066.01 . Of the existence of these there is no evidence.

The calculated positions of the Q lines are $Q 1 = 11377.66$, $Q 2 = 11377.64$, $Q 3 = 11375.10$, $Q 4 = 11373.54$, $Q 5 = 11373.31$, $Q 6 = 11379.7$. $Q 5$ and $Q 4$ are represented by $11373.41(3)$ P ab G, defects -0.10 and $+0.13$, the rest are absent from Foetker's and Gale, Monk and Lee's tables.

The 1 → 1 Band

\wedge	$R 0$	13663 3(0a) G, also $Q 4$ of the $0 \rightarrow 0$ band.	
+0.02	$P 2$	13484.3(1) G.	
\vee			
\wedge	$R 1$	13715.7(3) G, also $\uparrow 0 \rightarrow 2$, $4d^1\Sigma \rightarrow 2p^1\Pi$, $R 2$.	
+0.12	$P 3$	13420.2(1) G.	
\vee			
\wedge	$R 2$	13765.9(1a) G, also $1 \rightarrow 0$, $3p^2\Sigma \rightarrow 2s^2\Sigma$, $R 3$.	
0	$P 4$	(13356 57) ab G. A has 13357.45(1).	
\vee			
\wedge	$R 3$	13808 1(1) G, also $Q 4$ of the $2 \rightarrow 2$ band	
-0.04	$P 5$	13290 4(00a) G.	
\vee			
\wedge	$R 4$	(ca. 13846.93) ab G and A	\wedge $R 4$ 13836.39(1) A ab G.
0	$P 6$	(ca. 13235.00) ab G and A	or +0.14 \vee $P 6$ 13224.6(0) G, also $P 7$ of the $0 \rightarrow 0$ band
\vee			
\wedge	$R 5$	13878.7(1) G.	
0	$P 7$	13190.0(0a) G.	
\vee			
	$Q 1$	13601.7(4) G, also $1 \rightarrow 0$, $3p^2\Sigma \rightarrow 2s^2\Sigma$, $P 1$.	
	$Q 2$	13594.8(00) G	
	$Q 3$	13582 6(1) G, also $P 2$ of the $0 \rightarrow 0$ band.	
	$Q 4$	13570 29(1) A, may be part of 13571 0(0a) G which is also $3 \rightarrow 4$, $3p^2\Pi \rightarrow 2s^2\Sigma$, $P 2$.	
	$Q 5$	13555.54(4) A G has 13555 8(3) = $4 \rightarrow 5$, $3p^2\Pi \rightarrow 2s^2\Sigma$, $Q 1$.	
	$Q 6$	\uparrow 13548.8(0) G, also $4 \rightarrow 5$, $3p^2\Pi \rightarrow 2s^2\Sigma$, $Q 2$.	

The 1 → 0 Band

$R 0$	(15991.26) ab G, MB, T and D.
$P 2$	(15801.47) may be overshadowed by 15800.76 (10).
$R 1$	16042.20(rd) D ab G, defect -0.65 .
$P 3$	15726.90(0) G, defect $+0.01$.
$R 2$	10083.48(rd) D ab G, defect -0.41 .
$P 4$	(15646.33) ab G, MB, T and D.
$R 3$	(16114.82) ab G, MB, T and D.
$P 5$	(15580.39) ab G, MB, T and D.

TABLE I—(continued)

<i>R</i> 4	16136·37(<i>rd</i>) D ab G, defect + 0·32 from 1 → 1, <i>R</i> 4 = (13846·93).
<i>P</i> 6	(15466·89) ab G, MB, T and D.
or <i>R</i> 4	16126·19(10) G, mostly 1 → 1, 3 <i>p</i> ² <i>I</i> → 2 <i>s</i> ² <i>Σ</i> , <i>P</i> 3, defect - 0·04 from 1 → 1, <i>R</i> 4 = 13836·39(1)
<i>P</i> 6	(15456·49) ab G, MB, T and D.
<i>R</i> 5	16149·27(<i>rd</i>) D ab G, defect - 0·58.
<i>P</i> 7	(? 15370·09) ab G, MB, T and D but interval uncertain.
<i>Q</i> 1	15927·80(1) G, also 4 → 1, ¹ X → 2 <i>p</i> ¹ <i>Σ</i> , <i>R</i> 0 and ? 0 → 1, ¹ N → 2 <i>p</i> ¹ <i>I</i> , <i>Q</i> 1, defect - 0·26.
<i>Q</i> 2	(15911·97) ab G, MB, T and D.
<i>Q</i> 3	15889 31(0) MB ab G, also ? 1 → 4, ³ A → 2 <i>p</i> ³ <i>I</i> , <i>P</i> 3, defect zero.
<i>Q</i> 4	15859·93(3) MB, defect + 0·12, 15859 86(5) G is 3 → 3, 3 <i>p</i> ² <i>I</i> → 2 <i>s</i> ² <i>Σ</i> , <i>R</i> 0.
<i>Q</i> 5	15825·06(<i>r</i>) D ab G, defect + 0·47.

The 1 → 2 Band

The calculated *R* lines are *R* 0 = 11596·42, *R* 1 = 11664·14, *R* 2 = 11747·46, *R* 3 = 11867·57, *R* 4 = 11996·94 or 11986 40, *R* 5 = 12136 51. These are all absent from Poetker's and Gale, Monk and Lee's tables.

The calculated *P* lines are *P* 2 = 11465·86, *P* 3 = 11479 67, *P* 4 = 11506·58, *P* 5 = 11548·2, *P* 6 = 11631·14 or 11620·60, *P* 7 = 11728·71. These are represented by *P* 3 = 11479 49(1) P ab G, defect + 0·18, *P* 5 = 11547·95(1) P = 11547 9(1*α*) G, also 4 → 3, 3*p*²*Σ* → 2*s*²*Σ*, *P* 4, defect + 0·25, *P* 6 = ? 11631·94(1) P ab G, defect - 0·80 and *P* 7 = ? 11728·0(1) G, defect + 0·71 (the interval for this line is uncertain). The rest are absent from P's and G's tables.

The calculated *Q* lines are *Q* 1 = 11550·14, *Q* 2 = 11576·36, *Q* 3 = 11642·07, *Q* 4 = 11720·29, *Q* 5 = 11813·35. These are all absent from P's and G's tables.

The 1 → 3 Band

The calculated lines are *R* 0 = 9771·12, *R* 1 = 9854·50, *R* 2 = 9963·67, *R* 3 = 10100·69, *R* 4 = 10269·47 or 10258·93, *R* 5 = 10453 46, *P* 2 = 9682 07, *P* 3 = 9712·79, *P* 4 = 9779·11, *P* 5 = 9865·16, *Q* 1 = 9740·50, *Q* 2 = 9792 57, *Q* 3 = 9875 19, *Q* 4 = 9992·82, *Q* 5 = 10130 30. These are all beyond Gale, Monk and Lee's infra-red limit and are absent from Poetker's table except *Q* 5 = 10130·30(*h*) P, defect zero.

The 2 → 2 Band

^	<i>R</i> 0	(13697·44) ab G ab P ab A.
0	<i>P</i> 2	13566 9(00 <i>a</i>) G, also ? 0 → 1, ¹ N → 2 <i>p</i> ¹ <i>I</i> , <i>Q</i> 3.
^	<i>R</i> 1	13757·8(2) G, also ? 0 → 4, ³ H → 2 <i>p</i> ³ <i>I</i> , <i>Q</i> 4.
-0 01	<i>P</i> 3	13573 4(1) G.
^	<i>R</i> 2	13831·5(2 <i>a</i>) G, also <i>R</i> 1 of the 0 → 0 band.
-0 18	<i>P</i> 4	13590·8(0) G.
^	<i>R</i> 3	13931 0(5) G, also <i>R</i> 3 of the 0 → 0 band.
0 00	<i>P</i> 5	13611·6(0) G.
^		

TABLE I—(continued)

\wedge -0.10?	R 4	? 14047.6(1) G, also $2 \rightarrow 3$, $3d^3\Sigma \rightarrow 2p^3\Pi$, Q 3.
\vee	P 6	? 13681.9(1) G.
		Alternative.
	R 4	? 14064.7(0) G.
	P 6	? (13698.9) ab G and A.
\wedge ?	R 5	? 14174 1(0) G.
\vee	P 7	?? 13766.3(0) G.
		Alternative.
\wedge ?	R 5	? 14190.8(00) G.
\vee	P 7	? coincident with 13783.1(10) G, also $2 \rightarrow 1$, $1X \rightarrow 2p^1\Sigma$, P 2.
	Q 1	13661.1(1) G.
	Q 2	13682 3(0c) G, also $1 \rightarrow 2$, $3p^3\Pi \rightarrow 2s^3\Sigma$, P 4.
	Q 3	13740 9(1) G, also $0 \rightarrow 0$, $3^1O \rightarrow 2p^1\Pi$, P 2.
	Q 4	13808.1(1) G, also R 3 of the $1 \rightarrow 1$ band.
	Q 5	13884.1(0a) G, also ? R 2 of the $0 \rightarrow 0$ band.
<i>The 2 → 1 Band</i>		
	R 0	(15764.32) ab G, MB, T and D.
	P 2	(15585.34) ab G, MB, T and D.
	R 1	{ 15809.00(1) MB, defect + 0.36. { 15808 94(0h) G, defect + 0.40.
	P 3	15513 96(0) G (q) T, defect - 0.03.
	R 2	(15849 94) ab G, MB, T and D.
	P 4	(15440 79) ab G, MB, T and D.
	R 3	{ 15872 08(0) MB ? $4 \rightarrow 4$, $3d^3\Pi_s \rightarrow 2p^3\Pi$, P 4, defect + 0.55. { 15872 20(1) G ? $3 \rightarrow 3$, $3d^3\Sigma \rightarrow 2p^3\Pi$, Q 2, defect + 0.67.
	P 5	(15353.79), probably absent.
	R 4	15897 61(rd) D ab G, MB and T, defect - 0.02.
or	R 4	(15914.69) ab G, MB, T and D.
	P 6	(15285.68) ? mixed with 15285.11(2) G, also $5 \rightarrow 5$, $3p^3\Pi \rightarrow 2s^3\Sigma$, Q 3.
or	P 6	(15302.78) ab G, MB, T and D.
	R 5	15916.34(rd) D ab G, MB and T, also $1_s \rightarrow 4$, $3H \rightarrow 2p^3\Pi$, P 3, defect - 0.05.
or	R 5	15933.16(0a) G, also $4 \rightarrow 1$, $2s^1\Sigma \rightarrow 2p^1\Sigma$, R_2 , defect - 0.17.
	Q 1	15712.79(6) G, also $2 \rightarrow 2$, $3p^3\Pi \rightarrow 2s^3\Sigma$, P 5, defect - 0.13.
	Q 2	(15700 74) ab G, MB, T and D.
	Q 3	15681.58(rd*) D, defect - 0.15.
	Q 4	(15658.1) ab G, MB, T and D.
	Q 5	? (15626.29) ? mixed with 15626.94(1) G, also $2 \rightarrow 2$, $3p^3\Pi \rightarrow 2s^3\Sigma$, P 6.

TABLE I—(continued)

The $2 \rightarrow 0$ Band

$R\ 0$	(18092.28) ab G, MB, T and D.
$P\ 2$	17902.41 [—] MB, He^{++} , defect + 0.09.
$R\ 1$	(18135.20), may be mixed with 18134.00(1) G [—] MB, He^{++} , also $1 \rightarrow 3$, $\uparrow \rightarrow 2p^1\Sigma$, $P\ 4$ and $\uparrow 1 \rightarrow 3$, $^3A \rightarrow 2p^3\Pi$, $R\ 3$.
$P\ 3$	(17820.64) ab G, MB, T and D.
$R\ 2$	(18167.10) ab G, MB, T and D.
$P\ 4$	(17730.55), may be part of 17729.30(q) T.
$R\ 3$	18178.16(9) G, also $0 \rightarrow 0$, $4d^1\Sigma \rightarrow 2p^1\Pi$, $R\ 2$ and $0 \rightarrow 0$, $4^1O \rightarrow 2p^1\Pi$, $P\ 4$, defect - 0.08.
$P\ 5$	17623.32(1) MB ab G, $C^{++}He^{++}$, defect - 0.54.
$R\ 4$	18187.02(1h) G, also $\uparrow 2 \rightarrow 0$, $3^1O \rightarrow 2p^1\Pi$, $Q\ 1$, defect - 0.57.
or $R\ 4$	(18204.45) ab G, MB, T and D.
$P\ 6$	(17517.54) ab G, MB, T and D.
or $P\ 6$	(17534.64) ab G, MB, T and D.
$R\ 5$	\uparrow (18186.28) or \uparrow (18202.98) both ab G, MB, T and D
$Q\ 1$	18038.17(00) G, defect + 0.33.
$Q\ 2$	18018.24(1) G, also $0 \rightarrow 4$, $^1L \rightarrow 2p^1\Sigma$, $R\ 2$, defect - 0.34.
$Q\ 3$	\uparrow 17988.16(πd^3) D ab G, MB and T, defect - 0.02.
$Q\ 4$	(17947.85) ab G, MB, T and D.
$Q\ 5$	(17895.28) ab G, MB, T and D.

The $2 \rightarrow 3$ Band

The calculated positions of the lines are $R\ 0 = 11872.14$, $R\ 1 = 11948.16$, $R\ 2 = 12047.71$, $R\ 3 = 12164.12$, $R\ 4 = 12320.13$ or 12337.23 , $R\ 5 = \uparrow 12491.05$ or $\uparrow 12507.75$, $P\ 2 = 11783.11$, $P\ 3 = 11806.52$, $P\ 4 = 11863.33$, $P\ 5 = 11928.55$, $Q\ 1 = 11851.46$, $Q\ 2 = 11898.41$, $Q\ 3 = 11974.02$, $Q\ 4 = 12080.63$, $Q\ 5 = 12201.05$.

These are represented by $R\ 1 = 11948.6(3)$ G, also $2 \rightarrow 1$, $^1X \rightarrow 2p^1\Sigma$, $P\ 3$, defect - 0.44; $R\ 3 = 12164.89(0^*)$ P ab G, defect - 0.77, $R\ 4 = 12319.5(1c)$ G, also $0 \rightarrow 0$, $3d^1\Sigma \rightarrow 2p^1\Pi$, $Q\ 4$ and $2 \rightarrow 1$, $3p^3\Sigma \rightarrow 2s^3\Sigma$, $P\ 7$, defect - 0.63; $P\ 3 = \uparrow 11807.32(0)$ P, defect - 0.80; $P\ 5 = 11928.2(2)$ G, also $4 \rightarrow 3$, $3p^3\Sigma \rightarrow 2s^3\Sigma$, $R\ 1$, defect - 0.35, $Q\ 3 = 11973.5(10)$ G, also $1 \rightarrow 1$, $^1X \rightarrow 2p^1\Sigma$, $P\ 2$, defect + 0.52; $Q\ 5 = 12201.1(00)$ G, defect - 0.05.

It will be seen from Table I that the P branches of the bands are definitely weaker than either the Q or R branches. This is also a property of the bands of the $np^3\Pi \rightarrow 2s^3\Sigma$ systems. They show the usual alternation of intensity, the lines with K'' odd being the stronger. In fact the lines with K'' even are for the most part missing from the weaker non-diagonal bands. Another property which they share with the bands of the $np^3\Pi \rightarrow 2s^3\Sigma$, and also of the $nd^3\Pi \rightarrow 2p^3\Pi$, systems is the concentration of the strength along the diagonal axis.

There are a few points of detail in Table I which require to be mentioned. The strong line (it is probably the most intense unclassified line now left in the spectrum) $R3$ of the $0 \rightarrow 0$ band is given by Gale, Monk and Lee as 13931.0 wave numbers, intensity (5). This number gives a very bad combination with the $0 \rightarrow 0$, $P5$ line. The line 13931.0 also does duty as $R3$ of the $2 \rightarrow 2$ band, and in this band it gives the corresponding combination exactly. On the other hand, the same line is given by Allibone as 13930.58, intensity (5). This difference, 0.42 wave number, is very large and cannot be attributed to any known blends. It compares with an average residual difference between Allibone and Gale, Monk and Lee of 0.004 wave number. Allibone's value reduces the defect in the $0 \rightarrow 0$, $R3-P5$ difference from 0.45 to 0.03 wave number. It seems clear that $0 \rightarrow 0$, $R3$ must have been enhanced in Allibone's discharge and $2 \rightarrow 2$, $R3$ in Gale, Monk and Lee's. There are so many blends in these bands and so much doubt about some of the intensity estimates that I have been unable to decide whether there is a corresponding general enhancement of the $0 \rightarrow 0$ and $2 \rightarrow 2$ bands as a whole, but one cannot be sure there is not. There is also little doubt that 13931.0 is too strong either as $0 \rightarrow 0$, $R3$ alone or as $2 \rightarrow 2$, $R3$ alone. Most of the strength of 13831.5(2a) G is required by $0 \rightarrow 0$, $R1$ which shares this line with $2 \rightarrow 2$, $R2$.

In the $1 \rightarrow 0$ band the failure of $R3$ and $P5$ is curious. However, all the lines are very weak. In general where any real strength is assigned to lines of the non-diagonal bands it is safe to assume that it is due to overlapping by a line of some other system. The only reason I have mentioned the line 18134.00(1) G [—] MB, He^{++} in connexion with $2 \rightarrow 0$, $R1$ where the defect would appear to be prohibitive is that the two existing allocations do not account for the He^{++} feature which seems to be almost invariably associated with transitions to or from vibrationless levels. The line 17988.16 (rd^b) D is probably an unresolved blend of 17989.11(1) G and 17987.29(1) G . If so, $2 \rightarrow 0$, $Q3$ is absent.

The initial rotational intervals of the present bands have been got from the wave numbers of the various P , Q and R lines of the stronger diagonal bands using relations of the type

$$F'_e(K+1) - F'_e(K) = R(K) - R(K-1) + F''_e(K) - F''_e(K-1). \quad (1)$$

For the lower K values the R and P lines generally yield two independent determinations of the intervals. The values for the Q branches are each the result of a single determination. The level differences thus obtained are set out in Table II. The corresponding structure in $3p^3II$ is shown in Table III. In this table two sets of values are given at $v = 1$ of $3p^3II_g$. The first set is

obtained by interpolation, being the mean of the values at $v = 0$ and $v = 2$. The second is got from the Q lines of $1 \rightarrow 0$, $3p^3\Pi \rightarrow 2s^3\Sigma$ and the known final differences of $2s^3\Sigma$ by the usual method. It will be seen that the results of the two methods are practically identical. This is of some importance because for the higher K levels of $3p^3\Pi$, at $v = 1$ interpolation is the only method available as this vibrational level disappears for $K > 3$, presumably owing to a perturbation.

TABLE II—THE ROTATIONAL STRUCTURE OF THE VIBRATIONAL LEVELS OF THE UPPER STATE

<i>R, P Branches</i>					
K ↓	$v' = 0$	$v' = 1$	$v' = 2$		
6	a —————			or	
	339.20	↑ 317.98	304.84		304.44
5	s -----→ 50.39	588.34 -----→ ↑ 47.62	-----→ 47.13		-----→ 29.68
	288.81	↑ 270.36	257.71		274.76
4	a -----→ 54.11	-----→ ↑ 50.32	-----→ 58.48		-----→ 75.53
	234.70	220.04	199.23		199.23
3	s -----→ 54.61	-----→ 52.23	-----→ 40.86		-----→ 40.86
	179.09	167.81	158.37		158.37
2	a -----→ 56.47	-----→ 54.10	-----→ 52.06		-----→ 52.06
	122.62	113.71	106.31		106.31
1	s -----	-----	-----		-----
<i>Q Branches</i>					
K ↓	$v' = 0$	$v' = 1$	$v' = 2$		
6	s -----	-----			
	332.0	↑ 318.93			
5	a -----→ 45.9	-----→ 47.42 ↑	-----		
	286.10	271.51	254.18		
4	s -----→ 56.35	-----→ 52.29	-----→ 45.92		
	229.75	219.22	208.26		
3	a -----→ 54.53	-----→ 53.62	-----→ 49.84		
	175.22	165.60	158.42		
2	s -----→ 57.49	-----→ 54.84	-----→ 52.63		
	117.73	110.76	105.79		
1	a -----	-----	-----		

The rotational structure of the upper state of the present bands is very similar to that of $3p^3\Pi$. Indeed a great many of the corresponding level differences are almost identical. It does not attain the very unusual regularity, for a H₂ band system, shown by $3p^3\Pi$. Some of the smaller irregularities may be due to a lower accuracy in the data. Most of the data for

$3p^3\Pi$ are based on the means of a considerable number of independent determinations, and this wealth of material is not available in connexion with the present bands. Evidently there are some abnormalities. The $K = 5$, s level at $v' = 1$ shown in Table II is got from extrapolated lines which would make the structure regular and there is no evidence of its real existence. The failure of this level, i.e. the absence of the $R4$, $P6$ lines,

TABLE III—THE ROTATIONAL STRUCTURE OF THE
CORRESPONDING LEVELS OF $3p^3\Pi_b$

		<i>R, P Branches</i>			
K ↓		$v=0$	$v=1$, interpolated	$v=2$	
6	<i>a</i>	340 15	322 58	305 01	
5	<i>s</i>	-----→ 52 10 287 95	-----→ 48 96 273 62	-----→ 45 71 259 30	
4	<i>a</i>	-----→ 54 65 233 30	-----→ 51 58 222 04	-----→ 48 52 210 78	
3	<i>s</i>	-----→ 56 61 176 69	-----→ 53 67 168 37	-----→ 50 73 160 05	
2	<i>a</i>	-----→ 58 08 118 61	-----→ 55 26 113 11	-----→ 52 43 107 62	
1	<i>s</i>	-----	-----	-----	

		$3p^3\Pi_a$ (<i>Q Branches</i>)			
K ↓		$v=0$	$v=1^*$	$v=1^\dagger$	$v=2$
6	<i>s</i>	----- 339 18			
5	<i>a</i>	-----→ 52 45 286 63	----- 272 12	----- 271 96	----- 257 62
4	<i>s</i>	-----→ 54 71 231 92	-----→ 51 84 220 28	-----→ 51 74 220 22	-----→ 48 97 208 65
3	<i>a</i>	-----→ 56 41 175 51	-----→ 53 57 166 71	-----→ 53 59 166 63	-----→ 50 73 157 92
2	<i>s</i>	-----→ 57 71 117 80	-----→ 54 81 111 90	-----→ 54 88 111 75	-----→ 51 91 106 01
1	<i>a</i>	-----	-----	-----	-----

* Interpolated (mean of $v=0$ and $v=2$).

† Got from the Q lines of $1 \rightarrow 0$, $3p^3\Pi \rightarrow 2s^3\Sigma$ and the known final differences of $2s^3\Sigma$.

may be just due to weakness and have no particular significance. However, there is a pair of lines, given in Table I, which gives the combination, and which would place the level about 12 wave numbers lower; so there may be a perturbation. There is a definite perturbation at the $K = 4$ level at $v' = 2$, the upper level of the $R3$, $P5$ lines, which is too low. This makes

it difficult to locate the $K = 5$ and $K = 6$ levels with certainty. However, the pair of $2 \rightarrow 2$ lines which lead to the $K(5)-K(4)$ interval 257.71 at $v' = 2$ have several combinations with lines off the diagonal axis and are thus the more probable. They also lead to a less irregular rotational structure. Two possible sets of these levels are given in Table II corresponding to the two alternative sets of R, P lines in Table I. The irregularity in the Q levels at $v' = 0$ appears to occur at $K = 4$. If $Q 4$ were raised about 2 wave numbers the structure would be regular. The $Q 4$ line may be too weak to have been recorded but the α feature is evidence that $13663.3(0a)$ is more than one line.

The initial vibrational differences have been extracted from the lines of the diagonal bands by means of relations of the type

$$\Delta'[(v+1)-(v)](K) = \Delta''[(v+1)-(v)](K+1) + (v+1) \rightarrow (v+1), \\ P(K+1) - v \rightarrow v, P(K+1). \quad (2)$$

As in the case of the rotational intervals this gives, in general, two independent determinations of the differences for the R, P levels and only one for the Q levels. The values thus found are set out in the top half of Table IV. The two values in brackets at $K = 5$ depend upon lines whose wave numbers have been estimated from the band structure and not either upon real lines or upon lines whose wave numbers can be determined by established combinations. The alternative values given in the footnotes correspond to the alternative lines given in Table I. There is another possible pair of alternatives, not given, at $K' = 5$ of $3p^1\Pi_v$, $v = 1$. These use the $R 4, P 6$ lines at $v' = 1$ in Table I which give the combinations but make the structure irregular. The corresponding vibrational differences of $3p^3\Pi_{ab}$, taken from Richardson and Das (1929), are set out in the lower half of Table IV. It will be seen that the two sets of level differences are very much alike. The greater irregularity of the values for the upper state of the present band system is intimately connected with the irregularities in its rotational structure which have been noted already.

The evidence is thus steadily accumulating that the bands I am describing are $^1\Pi_u \rightarrow ^1\Sigma_g$ bands. The fact that the final state is $1s\sigma^2 2s\sigma^1 \Sigma_g$ has been established, so that we do not need to consider it further. The upper state must be a singlet (prefix 1) state and also an odd (suffix u) state otherwise it would not have transitions down to $2s^1 \Sigma_g$. Other facts which support its allocation as a $^1\Pi_u$ state are the presence of P, Q and R branches, the very close resemblance of its rotational and vibrational structures with those of the well-known $3p^3\Pi_{ab}$ state as well as the general resemblance of the whole band system with the $np^3\Pi_{ab} \rightarrow 2s^3 \Sigma_g$ systems. The fact that the $K' = 1$

level is present proves that the upper state is either a Π or a Σ state. The question of the existence of $P1$ lines which would come from a $K' = 0$ level is therefore of importance. If the existence of such lines could be established it would be necessary to interpret the R, P branches as a

TABLE IV
VIBRATIONAL DIFFERENCES $\Delta\nu'$ OF THE UPPER STATE

$K \rightarrow$	1	2	3	4	5	6
Mean (1-0), R, P	2219.01 9.04	2209.97 11.22	2198.75 14.49	2184.26 18.56	(2165.70)* 21.1	2144.6
(1-0), $Q(K)$	2224.04 6.87	2217.17 9.76	2207.41 10.66	2196.75 14.52	2182.23 13.14	? 2169.09
Mean (2-1), R, P	2101.03 7.34	2093.69 7.56	2084.13 20.72	2063.41 13.67	(2050.74)† 12.15	2037.59†
(2-1), $Q(K)$	2110.96 5.01	2105.95 7.12	2098.83 11.03	2087.80 17.05	2070.75	

* Alternative 2155.22.

† Alternative $K = 5$ (2066.79), $K = 6$, 2054.34.

VIBRATIONAL DIFFERENCES $\Delta\nu$ OF $3p^3\Pi_{ob}$

$K \rightarrow$	1	2	3	4
(1-0), $3p^3\Pi_b, R, P$	2239.45 8.22	2231.23 14.95	2216.28 19.49	2196.79
(1-0), $3p^3\Pi_a, Q(K)$	2240.21 5.98	2234.23 8.95	2225.28 11.84	2213.44
(2-1), $3p^3\Pi_b, R, P$	2116.08 2.71	2113.37 1.65	2111.72 3.22	2108.50
(1-0), $3p^3\Pi_a, Q(K)$	2114.84 5.76	2109.09 8.71	2100.38 11.38	2089.00

$^1\Sigma_u \rightarrow ^1\Sigma_g$ system and the Q branches as something else. At one time I thought that I had found such $P1$ lines but they have not been able to withstand a closer examination. They are near, but not near enough to their expected positions and tend to spoil rather than improve the regularity of structure already present. They are all wanted elsewhere and after allowing for the foreign intensity there is not enough left for them to function as $P1$ lines of the present system. This, then, rules out the possibility of $^1\Sigma_u \rightarrow ^1\Sigma_g$ so that

the upper state must be a ${}^1\Pi_u$ state. The similarity of its structure to that of $1s\sigma 3p\pi {}^2\Pi_u$ strongly suggests that it is the corresponding singlet state $1s\sigma 3p\pi {}^1\Pi_u$, making the band system $1s\sigma 3p\pi {}^1\Pi_u \rightarrow 1s\sigma 2s\sigma {}^1\Sigma_g$ or, in the contracted nomenclature which is adequate for states with only one electron excited, $3p {}^1\Pi_u \rightarrow 2s {}^1\Sigma_g$.

We have seen that it has been possible to identify the $3p {}^1\Pi_u$ state with the upper state, D, of Hopfield's absorption bands. If the upper state of the present band system is also $3p {}^1\Pi_u$ it should lie at the same height above the ground state of H_2 as the D state does. Data are available for testing this question. The lowest level $K = 1, v = 0$ of the upper state of the present band system will lie above the $K = 0, v = 0$ level of $2s {}^1\Sigma_g$ ($\equiv {}^1X$) by an amount equal to the wave number of the $0 \rightarrow 0, R0$ line of the bands. The $K = 0, v = 0$ level of $2s {}^1\Sigma_g$ will, in turn, lie above the $K = 1, v = 0$ level of $2p {}^1\Sigma_u$ by the wave number of the $0 \rightarrow 0, P1$ line of ${}^1X \rightarrow 2p {}^1\Sigma$. Finally the $K = 1, v = 0$ level of $2p {}^1\Sigma_u$ will lie above the $K = 0, v = 0$ level of the ground state $1s {}^1\Sigma_g$ by the wave number of the $0 \rightarrow 0, R0$ line of $2p {}^1\Sigma \rightarrow 1s {}^1\Sigma$. According to the foregoing analysis the $0 \rightarrow 0, R0$ line of the present system is 13772.1 wave numbers, Dieke's value of $0 \rightarrow 0, P1, {}^1X \rightarrow 2p {}^1\Sigma$ is 8922.38 and the value of $0 \rightarrow 0, R0, 2p {}^1\Sigma \rightarrow 1s {}^1\Sigma$ given by Richardson (1934, p. 156) is 90233.5. I have checked the last by making three other independent determinations of it. It is difficult to assess the relative accuracy of the data used in these computations but the new values alone would not seem to afford any compelling reason for changing the old one. The sum of the wave numbers of these three lines is 112927.98, or say 112928, wave numbers. This then is the height of the lowest ($K = 1, v = 0$) level of the upper state of the present system above the lowest ($K = 0, v = 0$) level of the ground state of H_2 .

According to Beutler, Deubner and Jünger (1935) the $3 \leftarrow 0, R0$ line of ${}^1D \leftarrow 1s {}^1\Sigma$ is 119227 wave numbers. Owing to the occurrence of predissociation only the vibrational levels of 1D with $v > 2$ can be detected in absorption. It is only possible to get from their data for the higher vibrational levels down to the $v = 0$ level of 1D by a rather hazardous extrapolation. On the other hand I have, so far, only been able to find the $v' = 0, 1$ and 2 levels of the present system. According to Beutler, Deubner and Jünger the $v' = 3$ (and higher) levels should not be found in emission. If the $v' = 3$ level does occur it is not easy to find and I am disposed to accept their contention provisionally pending a further examination of the spectrum. It is thus necessary to extrapolate from the $v' = 0, 1$ and 2 levels to the next higher $v' = 3$ level if we are to make a connexion with the data for the D state. Turning to Table IV we see that the $K = 1$ initial differences for the R, P

branches are at $\Delta v' = (1-0)$, 2219.01 and at $\Delta v' = (2-1)$, 2101.03. The difference 2219.01 - 2101.03 is 117.98. Assuming this difference to be constant as v increases by unity, as it usually is to a close approximation apart from perturbations, the value at $\Delta v' = (3-2)$ comes to 2101.03 - 117.98 = 1983.05. From the trend of the data for both states this should be about 10 wave numbers below the $\Delta v' = (3-2)$ value for $3p^3\Pi_b$, which is 1994.42, so that there is little doubt that the value 1983.05 is very nearly correct.

Using these vibrational data and assuming that the upper levels of the present system are the continuation at $v' = 2, 1$ and 0, starting from the measured value 119227 of $3 \rightarrow 0, {}^1D \rightarrow 1s^1\Sigma, R0$, we have, $2 \rightarrow 0, {}^1D \rightarrow 1s^1\Sigma, R0 = 119227 - 1983.05 = 117243.96$, $1 \rightarrow 0, {}^1D \rightarrow 1s^1\Sigma, R0 = 115142.92$ and $0 \rightarrow 0, {}^1D \rightarrow 1s^1\Sigma, R0 = 112923.91$ or say 112924 wave numbers. This is in excellent agreement with the value 112928 wave numbers got by adding together the cascade of three transitions starting from the $K' = 1, v' = 0$ level of the present band system and ending on the $K = 0, v = 0$ level of the ground state of H_2 . This, then, establishes the identity of the upper state of this system with 1D and with $1s\sigma 3p\pi^1\Pi_u$.

The other properties of the upper state and of the band system confirm this identification. Taking the ground state of the molecular ion $1s^2\Sigma_g$ to lie 124429 above that of H_2 and subtracting the mean 112926 of the two estimates of the wave number of $0 \rightarrow 0, R0$ of $3p^1\Pi \rightarrow 1s^1\Sigma$ I find that the $K = 1, v = 0$ level of $3p^1\Pi$ lies 11503 below $1s^2\Sigma_g$ and allowing 60 for the rotational energy $F'_0(1) - F'_0(0)$ this makes the ν_e of $3p^1\Pi = 11563$. This corresponds to a Rydberg denominator of 3.0803.* The state $1s\sigma 3p\pi^1\Pi$ is thus just above $1s\sigma 3s\sigma^1\Sigma$ which is the highest of the other fourteen known $1s\sigma 3(s, p \text{ or } d)\Sigma, \Pi$ or Δ states of H_2 . For $1s\sigma 3s\sigma^1\Sigma, \nu_e = 11613$ and the denominator is about 3.07. The comparatively high position of $1s\sigma 3p\pi^1\Pi$ is in conformity with that of $1s\sigma 2p\pi^1\Pi$, the next and lowest state of the same Rydberg series terminating at the ion $1s^2\Sigma$, for which the denominator is 2.0751.* Just as $2p^1\Pi$ and $2s^1\Sigma$ are the two highest of the H_2 states with only one electron excited and that to principal quantum number 2, so $3p^1\Pi$ and $3s^1\Sigma$ are the two highest with only one electron excited and that to principal quantum number 3.

On account of the irregularities in the rotational structure it is not possible to evaluate the B'_v 's very accurately. By a graphical comparison of the structure of the three lowest levels with those of the very regular $1s\sigma 3p^3\Pi$ I have arrived at the following values: For $1s\sigma 3p^1\Pi_b$ (R, P branches) B_0 = about 29.4, $B_1 = 27.9$, and $B_2 = 26.8$; for $1s\sigma 3p^1\Pi_a$

* These Rydberg denominators are in terms of $R_{H_2} = 109710 \text{ cm.}^{-1}$. To get the denominators in terms of the more usual $R_H = 109678.3$, divide by 1.000145.

(Q branches) $B_0 = 29.5$, $B_1 = 28.2$, and $B_2 = 27.2$. The values given in Richardson (1934) for $3p^3\Pi_b$ are $B_0 = 29.825$ and $B_2 = 27.105$ wave numbers. There is evidently very little difference between these constants for the singlet and the triplet states. The value for $3p^1\Pi$ of α , the quantity in the equation $B_v = B_0 - \alpha_v$, is not very far from 1.2 on the average. In Richardson (1934) α for $3p^3\Pi_b$ is given as 1.36 and for $3p^3\Pi_a$ as 1.30, so that there is not very much difference here. In the singlets also α looks to be smaller for Π_a than for Π_b .

Dieke's values of the B_v 's for the lower state $1s\sigma 2s\sigma^1\Sigma_g$ are $B_0 = 31.770$, $B_1 = 29.952$ and $B_2 = 24.23$. The closeness of these values to the B_v 's of the upper state $1s\sigma 3p\pi^1\Pi_u$ accounts for the concentration of the intensity in the bands which lie along the diagonal sequence and is thus a confirmation of the numbering which I have assumed for the vibrational levels of this state.

The vibrational constants of $1s\sigma 3p\pi^1\Pi$ cannot be determined very exactly from the vibrational differences of the $v' = 0, 1$ and 2 levels of the present system. There is, however, little doubt that they are close to the following values: For $3p^1\Pi_b$, $\omega_0 = 2282.5$ and $x\omega_0 = 61.2$, for $3p^1\Pi_a$ $\omega_0 = 2285.1$ and $x\omega_0 = 58.8$. Whether the difference between $3p^1\Pi_b$ and $3p^1\Pi_a$ represents anything real or not is doubtful. If we take the means we have: For $3p^1\Pi_{ab}$, $\omega_0 = 2283.8$ and $x\omega_0 = 60.0$. From my re-interpretation of Hopfield's data I found (Richardson 1934) $\omega_0 = 2257$ and $x\omega_0 = 59.5$. The more recent values got from absorption data by Bentler and others (1935) are $\omega_0 = 2255.33$ and $x\omega_0 = 60.5$. The difference of 28 wave numbers in ω_0 between the values got from the two lowest vibrational levels and those got from the levels with $v' > 2$ seems to represent something real. It looks as though one effect of the perturbation was to loosen the vibrational structure of all the predissociating levels.

The corresponding quantities for $3p^3\Pi$ given in Richardson (1934) are: For $3p^3\Pi_b$ and $3p^3\Pi_a$ $\omega_0 = 2307.5$, and for $3p^3\Pi_a$ $x\omega_0 = 65.0$. The vibrational structure of the triplets is thus a little, but not much, tighter than that of the singlets.

3—THE v' PROGRESSION, $x^1\Sigma_u \rightarrow 2s^1\Sigma_g$

I shall now describe a v' progression ending on $1s\sigma 2s\sigma^1\Sigma_g$ and coming from an upper state which I call provisionally $x^1\Sigma_u$. What strength there is lies in the $v' \rightarrow 0$ band which is closely interwoven with the $0 \rightarrow 0$ band of $3p^1\Pi \rightarrow 2s^1\Sigma$. In fact the existence of this band caused me a great deal of trouble in working out the details of the $0 \rightarrow 0$ band of $3p^1\Pi \rightarrow 2s^1\Sigma$. It

appears to have R , P but no Q branches and there is a $P\ 1$ line in the $v' \rightarrow 0$ band. As it goes down to $2s^1\Sigma_g$ the upper state must be an odd singlet state. If the $P\ 1$ line belongs to it the $K' = 0$ level must exist, so that it is a Σ state. I therefore indicate it by $x^1\Sigma_u$; what further interpretation should be put on the symbol x will require a fuller investigation of this section of the H_2 spectrum.

The lines are given in Table V in the same form as those of $3p^1\Pi \rightarrow 2s^1\Sigma$ in Table I. The $R\ 1$ - $P\ 3$ combination of $v' \rightarrow 0$ has to do duty as the same combination in $0 \rightarrow 0$ of $3p^1\Pi \rightarrow 2s^1\Sigma$. It probably has enough strength for both positions. $R\ 1$ is also $R\ 2$ of $3p^1\Pi \rightarrow 2s^1\Sigma$. The fact that the intensity estimate of the line is marked a by Gale, Monk and Lee is an indication that it is at least double. The $R\ 2$ line is given by Allibone but not by Gale, Monk and Lee. Allibone's intensity estimates tend to run higher than theirs; so that his value (2) would probably have to be reduced on their scale. The $R\ 3$ line is almost certainly a good deal too strong as $Q\ 1$ in the triplet band. The remaining coincidences are weak lines in the present bands and do not require to take much strength away from the other lines of the blends.

TABLE V—THE LINES OF $x^1\Sigma_u \rightarrow 2s^1\Sigma_g$

<i>The $v' \rightarrow 0$ Band</i>	
	$P\ 1$ 13642.3(4), also $0 \rightarrow 0$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $Q\ 5$.
\wedge	$R\ 0$ 13770.5(1) G.
-0.06	$P\ 2$ 13580.8(0) G.
\vee	
\wedge	$R\ 1$ 13831.5(2a) G, also $0 \rightarrow 0$, $R\ 1$ and $2 \rightarrow 2$, $R\ 2$ of $3p^1\Pi \rightarrow 2s^1\Sigma$.
-0.05	$P\ 3$ 13517.0(2) G, also $0 \rightarrow 0$, $P\ 3$ of $3p^1\Pi \rightarrow 2s^1\Sigma$.
\vee	
\wedge	$R\ 2$ 13888.93(2) A ab G.
-0.24	$P\ 4$ 13452.7(1) G (U) A, also $3 \rightarrow 4$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $P\ 4$.
\vee	
\wedge	$R\ 3$ 13945.5(8) G, also $1 \rightarrow 2$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $Q\ 1$.
+0.05	$P\ 5$ 13390.0(0) G.
\vee	
\wedge	$R\ 4$ 14003.1(1) G, also $2 \rightarrow 3$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $R\ 3$ and $0 \rightarrow 0$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $R\ 5$.
-0.11	$P\ 6$ 13333.4(00) G, also $1 \rightarrow 7$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $Q\ 3$.
\vee	
\wedge	$R\ 5$ 14064.7(0) G.
+0.10	$P\ 7$ 13286.0(9) G, mostly $2 \rightarrow 0$, $2s^1\Sigma \rightarrow 2p^1\Sigma$, $P\ 2$.
\vee	
<i>The $v' \rightarrow 1$ Band</i>	
	$R\ 1$ 11505.77(0) P ab G, defect -0.07, also $0 \rightarrow 1$, $3p^1\Pi \rightarrow 2s^1\Sigma$, $R\ 1$.
	$R\ 5$? 11796.44(0*) P ab G, part of, defect -0.73.
<i>The $v' \rightarrow 2$ Band</i>	
	$R\ 3$? 9697.66(0) Dicke, part of, defect +0.60.

Allowing for the blends there is evidence of the usual alternation of intensity in the band lines, those with K'' odd being the stronger. This is also supported by the three lines which appear for $v'' > 0$. All of these occur at odd values of K'' . For the $R1$ line of the $v' \rightarrow 1$ band the combination is good. The $R3$ line is absent and for the $R5$ line the defect is 0.73 wave numbers. However, this line is marked (0*) in Poetker and the star denotes a line which is exceptionally difficult to measure. Such lines are usually blends. In the $v' \rightarrow 2$ band, I think the existence of a line, measured provisionally as 9697.66 wave numbers, intensity (0), by Dieke, with a defect of 0.60 wave number is some support for the existence of the $R3$ line. The final corrections to these lines far down in the photographic infra-red are at present uncertain, although it is probable that they are small; there are only about six known lines per hundred wave numbers in this region and, in any event, the line might be a blend. The P branches appear to be absent both at $v'' = 1$ and at $v'' = 2$.

The rotational structure of the upper state, got in the same way as that of the $3p^1\Pi$, is shown in Table VI. This structure is a peculiar one for a state of H_2 . It is characterized by slowly and uniformly increasing second

TABLE VI—ROTATIONAL STRUCTURE OF $x^1\Sigma_u \rightarrow 2s^1\Sigma_g$ AT $v' = ?0$

K ↓			
6	a	—————	
		368 70	
5	s	-----→ 62.67	
		306.03	> 1.18
4	a	————→ 61.49	
		244.54	> 0.90
3	s	-----→ 60.59	
		183.95	> 0.96
2	a	————→ 59.63	
		124.32	> 1.03
1	s	-----→ 58.60	
		65.72	
0	a	—————	

differences as K' increases. In fact none of the third differences shown in Table VI differs from the mean value 1.02 by more than 0.16 wave number. This is within the degree of inaccuracy which could come solely from the errors of measurement of the lines. It looks as though a slow and uniform uncoupling were taking place. It should perhaps be mentioned that there is an equally good combination at $R5 = 14069.1(0)$ G, also $0 \rightarrow 1$ of $3p^1\Pi \rightarrow 2s^1\Sigma$, $Q3$. $P7 = 13290.4(00a)$ G instead of the lines in Table V. This would, of

course, if these were the band lines, spoil the regularity at the highest difference.

The B'_v of $x^1\Sigma_u$ is evidently close to B_0 of $3p^1\Pi$ and so to B_0 of the final level, that of $v = 0$ of $2s^1\Sigma$. The closeness of these B 's strongly suggests that the value of v' is 0 and so puts the band with the strength in it on the diagonal axis at $0 \rightarrow 0$. In fact this identification of v' could be made to look plausible also on other grounds. If $v' = 0$ then the v_e of $x^1\Sigma_u$ is very close to that of $3p^1\Pi$ and a possible identification would be with $1s\sigma 4p\sigma^1\Sigma_u$ which should lie lower than $1s\sigma 4p\pi^1\Pi_u$. However, it seems idle to speculate further about such possible identifications till more is known about the spectrum.

4—OTHER BAND SYSTEMS

None of the bands discussed in this paper is identical with those referred to in my letter to *Nature* (Richardson 1935). In particular the bands mentioned there as coming from a state called 1D , and identified with Hopfield's D state, and ending on 1X ($2s^1\Sigma$) are different from the $3p^1\Pi \rightarrow 2s^1\Sigma$ bands and their upper state is different from the D state. Possibly they may be part of the $3p^1\Sigma \rightarrow 2s^1\Sigma$ system. Mr Newbold and I are continuing the examination of these bands and a number of fragments found since that date and hope to report further progress later.

In conclusion I should like to thank Professor Dieke of Johns Hopkins University for keeping me informed of his progress with the $2s^1\Sigma \rightarrow 2p^1\Sigma$ system ahead of publication, for a list of lines beyond the limits of any published tables and for other valued information touching the H_2 spectrum, and Mr Newbold for pointing out an error which was affecting a few of the weaker rotational intervals.

SUMMARY

This paper deals mainly with an account of the first band system ending on the (lower) state $1s\sigma 2s\sigma^1\Sigma_g$ (1X) of the H_2 spectrum to be analysed. The systems ending on 1X are of importance for a variety of reasons. They form the only missing link hitherto existing in the lower states of this spectrum. The final state is the deepest state which has been found up to the present to give transitions down to any state above the ground state of H_2 and is at present also the highest state to which transitions down, with emission of radiation, have been observed. The upper state is shown to be $1s\sigma 3p\pi^1\Pi_u$ and the component $1s\sigma 3p\pi^1\Pi_g$ is shown to be identical with Hopfield's

D state. The structure of the upper state is given and its usual constants deduced.

The paper also contains an account of a progression of another state which goes down to the same end state.

REFERENCES

- Allibone 1926 *Proc. Roy. Soc. A*, **112**, 196-212.
 Beutler and Junger 1936 *Z. Phys.* **101**, 285-303.
 Beutler, Deubner and Junger 1935 *Z. Phys.* **98**, 181-97.
 Doodhar 1926 *Proc. Roy. Soc. A*, **113**, 420-32.
 Dieke 1936 *Phys. Rev.* **50**, 797-805.
 Funksburg and Meeke 1929 *Z. Phys.* **54**, 597-631.
 Gale, Monk and Lee 1928 *Astrophys. J.* **67**, 89-113.
 Merton and Barratt 1922 *Philos. Trans. A*, **222**, 369-400.
 Poetker 1927 *Phys. Rev.* **30**, 418-28.
 Richardson 1934 "Molecular Hydrogen and its Spectrum." New Haven: Yale University Press.
 — 1935 *Nature, Lond*, **135**, 99-100.
 Richardson and Das 1929 *Proc. Roy. Soc. A*, **122**, 688-718.
 Richardson and Davidson 1929a *Proc. Roy. Soc. A*, **123**, 466-88.
 — 1929b *Proc. Roy. Soc. A*, **124**, 69-88.
 Richardson, Davidson, Marsden and Evans 1933 *Proc. Roy. Soc. A*, **142**, 63-76.
 Tanaka 1925 *Proc. Roy. Soc. A*, **108**, 592-606.

The Surface Layer of Polished Silica and Glass with Further Studies on Optical Contact

BY LORD RAYLEIGH, F.R.S.

(Received 5 March 1937)

[Plate 4]

1—INTRODUCTORY

In former work (Rayleigh 1936) it was noticed that when specimens of glass or silica were placed in optical contact, and the small residual reflexion from the interface was observed, the intensity of this reflexion varied in different specimens in a way that seemed difficult to explain. The work was begun with the idea of obtaining an estimate of the distance by the measurement of this reflexion, applying the ordinary formula for the reflexion from a thin plate. But the distance calculated in this way

(10-30 Å) was variable, and seemed too large to appear plausible, considering that the plates were close enough to exert very large attractive forces on one another.

2—TESTING REFLEXION OF SILICA IN AN IMMERSION FLUID

In pursuing the matter, the idea suggested itself that surfaces prepared in various ways could be tested far more easily and conveniently by immersing them in a liquid of equal index than by putting them into optical contact with similar plates. The need for flatness, and the somewhat difficult and uncertain technique of "contacting" are in this way avoided.

It was decided to begin with surfaces of fused silica, because of the considerable interface reflexion which had been observed with this material (Rayleigh 1936, p. 339) and because of its chemical simplicity and definiteness.

The photometric arrangements were the same as before. Two similar opal-bulbed lamps were used, verified to be equally bright. One was reflected in the immersed specimen, and the other in a single polished surface of silica in air (3.5 % reflexion). The latter was much the brighter, and was cut down by a revolving sector of variable aperture, before being brought to a Lummer cube for comparison with the other. The circumference of the sector disk was 1000 mm., and the value of the opening is always given in mm. 1 mm. sector opening is equivalent to $10^{-3} \times 3.5 \times 10^{-3}$, or 3.5×10^{-6} of the intensity of the incident light.

The liquid chosen was carbon tetrachloride, which is slightly more refractive than silica. Ether was added to adjust the index. The adjustment was made by immersing a prism of fused silica in a tank of the liquid, and observing a distant slit with a small Galilean telescope through the nearly direct vision prism thus constituted. The slit was seen simultaneously undispersed by light passing over the top of the prism, or past its edge. Ether was added until the refraction was equalized for the green of the spectrum (about the position of the mercury line $\lambda 5461$). The composition arrived at was carbon tetrachloride 100 c.c., ether 4 c.c., and was adopted as standard.

It is to be noticed that the liquid thus prepared has the same index as the *interior* of the material. If the surface layer is identical with the interior, reflexion should be evanescent. If by any treatment the surface layer is made different, reflexion should come into evidence.

The fired surface of sheet silica as produced by the Thermal Syndicate of Wallsend is wavy and very much inferior in flatness to the fired surface of

ordinary sheet glass. This makes the reflexion difficult to examine satisfactorily, but it was readily proved that the reflexion in the immersion fluid was very small. Sector 0.5 mm. or less, intensity referred to incident light 1.75×10^{-5} or less.

Commercially polished silica sheets as obtained from the maker showed a similar low reflexion, but there was a lack of uniformity. It was noticed that the edges of a small piece which had been polished by itself were brighter than the middle; and there were irregularities even in the middle of large pieces, some regions being notably brighter than others.

The darkest parts were balanced by a sector aperture of 1 mm. Since the piece was parallel, half of this may be credited to each side, so that the equivalent for one surface would be 0.5 mm. (intensity referred to incident light 1.75×10^{-5}), as in the case of the fused surface.

3—CONDITIONS FOR PRODUCING A HIGHLY REFLECTING LAYER ON SILICA

Repolishing the surface at once produced striking and indeed astonishing results. The early attempts to repolish by any kind of amateur process that was tried always increased the (immersed) reflexion, usually making it many times brighter than before. This was the effect of polishing by hand with rouge used on a lump of pitch, whether the pitch was hard or soft, and whether the rouge was wet or comparatively dry. A wax polisher (rouge incorporated in paraffin wax) behaved similarly. A polisher of paper used dry with rouge also gave the bright reflexion, but by further working in damp weather, the reflecting power developed in this way diminished again and the original condition was nearly recovered. By using a polisher previously dried in a desiccator, and keeping the specimen slightly warm, and therefore dry, the reflecting power could be increased. By breathing on the silica so as to make it slightly damp the polisher was made to drag perceptibly more than before, and the reflecting power diminished.

Further tests with pitch polishers led to a similar result. Using a power driven pitch polisher very wet, and pressing lightly, the reflecting power was increased. Allowing it to become nearly dry, so that it dragged heavily, and tended to squeak, the reflecting power was diminished. This no doubt approximates to the condition under which the commercial polishing is done. The heavy dragging of the polisher indicates rapid removal of material, and effective working from the optician's point of view, when it is desired to pass from a fine ground surface to a polished one.

In order to prove more definitely that rapid removal of material was the

condition for leaving the surface with a low reflecting power, the loss of weight was tested as the process went on. A cloth polishing wheel was used, consisting of 6 in. disks of cloth assembled on a spindle. The edge was charged with wet rouge, and the surface of the specimen held against it. A wedge of silica was used, measuring 4×4 cm., only one face being treated and examined. After a period of treatment the wedge was cleaned, the loss of weight determined, and the reflecting power in liquid observed. When this had been tried with the wheel quite wet, no more water was added, but the process was repeated as the wheel became progressively drier.

In the following series of tests (Table I), each treatment lasted 5 min. The sector aperture (mm.) is given at the beginning and end of the treatment, and also the loss of weight in milligrams, when any could be detected.

TABLE I

Condition of wheel	Sector mm		Loss of weight mg.
	Initial	Final	
Kept quite wet	14	20	0.0
"	20	15	0.0
Drier (no more water)	15	10	0.0
"	10	5	0.0
"	5	3	0.05
"	3	1	0.15
"	1	5	0.15
"	5	8	0.10
"	8	11	0.00
Dry	11	11	0.00
Dry	11	11	0.00

It will be observed that the intensity of reflexion goes through a minimum as the polisher dries up, indicating that a certain low degree of moisture is the optimum for getting low reflexion. It is believed that a certain low degree of moisture is also the optimum for getting rapid polishing, i.e. rapid removal of substance. The changes of weight, though too small to be very satisfactory, on the whole bear this out. The reflexion is at a minimum when the material is being rapidly removed.

To increase the reflexion quickly to the utmost, I have succeeded best with a fine abrasive applied dry on a felt wheel. Fine abrasives such as carborundum or "abradum"* used dry on felt do not grind a glass or silica surface, though they do grind when used wet between glass and glass, or between glass and metal. Carborundum or "abradum" is definitely more

* Made by the County Chemical Works, Birmingham for finishing fine steel work such as gauges.

efficient in increasing the reflexion of a silica surface than rouge or putty powder used in the same way. The felt wheel without powder has practically no effect. "Abradum" is preferred to carborundum, as it seems to have less tendency to produce scratches. By about half an hour's treatment the sector value can be brought up to 60 mm. (intensity 2.1×10^{-3}) which seems to be about the limit easily attained, though small areas have on occasion been got up to 80 mm. (intensity 2.8×10^{-3}).

It appears that the process of increasing the reflecting power is not attended with attrition of material. So far as this occurs it is to the bad, and results in removing the modified material, and exposing unaltered silica.

An efficient polishing process which removes material rapidly, leaves a surface of low reflecting power, similar in character to the interior. If the polishing process is not rapid, the result is a kind of burnishing, which modifies the surface without removing anything, and gives it considerably increased reflecting power.

4—PROPERTIES OF THE REFLECTING LAYER ON SILICA

The highly reflecting layer is unaffected by lapse of time, by vigorous rubbing with a wet or dry cloth, by heating to redness in the blowpipe, or by prolonged treatment with hot sulphuric or chromic acid. On the other hand it is at once removed by treatment with dilute hydrofluoric acid and the low reflecting power restored. This has been found to be the method best adapted for getting the reflexion down to a minimum, for it removes the highly reflecting layer without any tendency to manufacture more of it in the process of removing what is already there.

By dipping a commercially polished silica wedge for a minute or two in dilute hydrofluoric acid (2% of the commercial acid) and then washing and wiping, the polish of the surface is not found to be markedly impaired, but on examining the immersed reflexion this is found to have gone down very conspicuously. The following (Table II) is a record of an experiment of this kind.

TABLE II

Time in hydrofluoric acid min.	Sector aperture mm.	Intensity (incident light = 1)
0	1	3.5×10^{-4}
1	0.25	0.9×10^{-4}
2	0.1 (roughly estimated)	3.5×10^{-4}

In the final condition the image was nearly invisible and the limit of the existing photometric method practically reached. I have not pursued this point further but it would perhaps be feasible to do so.

The strong reflecting power obtained by the methods already described can also be removed by 8% hydrofluoric acid (see below pp. 518-519), but some deterioration of the surface results.

Coming now to the optical properties, it will be noticed that the normal surface is optically like the interior, since its reflexion is annulled by a liquid of the same index as the body of the material. The modified layer differs from this, and in investigating its properties we are hampered by the circumstance that such modified material is not available in bulk for optical examination. All that we can do is to make observations of reflexion from the surface.

One important point is to test the normal reflexion in various liquids, and to find what index the liquid must have in order to reduce the reflexion to a minimum. For this purpose a silica wedge of 2° angle was used, with one surface modified. The photometric arrangement was that already described. The reflexion from either surface could be brought into view, and compared with the reflexion of a single silica surface in air. The tests were made with green light, using a green glass over the eye. In this way the embarrassment due to chromatic effects was avoided.

The liquids used were, according to the number in the first column of Table III:

- 1, Carbon tetrachloride with ether.
- 2, 3, Carbon tetrachloride with benzene—various compositions.
- 4, Benzene.
- 5-13, Benzene with carbon disulphide, various compositions.
- 14, Carbon disulphide.

Six different silica surfaces were examined. *a* was as received from the optician. It would have been better to use a surface treated with hydrofluoric acid, but this was not done, and it hardly seemed that time would be well spent in going back to it. *b*, *c* and *d* were successive stages of treatment of a single surface on the felt wheel with "abradum".

e is another surface which had been brought straight up to high reflexion in liquid No. 1. *f* represents a special effort to get the highest possible. The brightest part of the surface was observed, the photometric field being stopped down.

The observation giving minimum reflexion is in heavy type in each case. This of course indicates what liquid has about the same index as the surface, and it will be noticed how the index of the surface goes up as treatment



FIG. 2



FIG. 3

FIGS. 2 and 3, Plate 4, illustrate the methods of § 8, though the photographs reproduced are not those of the actual quantitative experiments described in the text. Fig. 2 shows the immersed reflexion from a strip of treated silica, with strips dipped for varying periods in hydrofluoric acid. In the darkest of them the modified layer has been virtually removed. Fig. 3 represents the dislocation of interference fringes across a strip eroded by the acid.

TABLE III

Sector values, mm. of opening

L.	Surface ... o. Index	Slides										Crown glass	Flint glass
		Unmodi- fied a	Modi- fied b	Modi- fied c	Modi- fied d	Modi- fied e	Modi- fied f	Crystal quartz		Crown glass			
								Original	Modified	Original	Modified		
1	1.461	1.2 B	5 N	12 N	20 N	60 N	80 N	25 N	25 N	12 N	25 N	—	—
2	1.481	—	4 N	8 N	17 N	—	—	—	—	—	—	—	—
3	1.489	—	4 N	8 N	15 N	—	—	—	—	—	—	—	—
4	1.507	4 B (?)	4 N	7 N	11 N	30 N	—	4 R	6 R (?)	12 R	—	—	—
5	1.525	10 B (?)	5 N	7 N	9 N	20 R	—	14 RR	2.1 R	9.25 N	3.1 R	—	—
6	1.529	—	5 N	7 N	9 N	—	—	—	—	—	—	14 N	—
7	1.541	15 B (?)	—	—	—	15 R	—	0.1 RR	0.65 R	1.2 BB	0.9 R	—	—
8	1.546	17 B (?)	8 N	9 N	7 R	13 R	—	0.0 RR	0.1 R	2.1 B	0.65 R	—	—
9	1.553	20 B (?)	10 N	10 N	8 R	10 R	—	0.25 BB	0.25 R (*)	2.1 B	0.25 R	6 R	—
0	1.565	27 B (?)	15 N	13 N	8 R	7 R	8 R	1.3 BB	0.65 B	6 B	1.0 B	3.3 R	—
1	1.592	40 B (?)	18 N	17 N	9 R	5 R	4 R	4.0 R	2.1 B	12 B	2.6 B	3.0 R	—
2	1.596	50 N	22 N	21 N	11 R	3 RR	4 RR	9.0 B	6.0 B	15 B	7.0 B	1.5 R (—
3	1.616	60 N	34 N	26 N	18 R	5 RR	3.3 RR	17.0 B	14.0 B	25 B	14 B	3.3 BB	—
4	1.635	75 N	44 N	36 N	42 R	7 RR	4 RR	22 B	18 B	37 N	18 N	4.5 B	—

proceeds. Initially it is the same as that of the body of the silica, but ultimately it may reach the surprisingly high value of 1.6. This is somewhat higher than the index of ordinary light flint glass, as used for telescopic objectives and also for table glass ware.

In the case of the modified surfaces we have to deal with a layer of modified material superposed on the normal interior. Reflexion will occur at the outside surface and also at the transition. The former can alone be annulled by the liquid, and the latter remains over and accounts for the residual intensity at minimum. In the case of the surface e this is 1.05×10^{-4} of the incident light.

This intensity, which, on the view here taken is *internal* to the solid, will remain whatever fluid is used. If we use a fluid of index 1.46, equal to that of the body of the silica, the observed reflexion of surface e is 2.1×10^{-3} . If the intensity due to internal reflexion is subtracted the residue is 2.0×10^{-3} , for the external reflexion, which is 19 times as great as the internal reflexion. Yet the modified surface layer is sandwiched in between two layers of the same index (one fluid, one solid) which under idealized conditions should give equal reflexions. It can only be concluded that the small value of the internal reflexion is due to the lack of abruptness in the transition from modified to unmodified material, contrasting with the abrupt transition between the polished external surface and the liquid.

Given the relative index, the reflexion for any liquid other than that which reduces the reflexion to a minimum is calculable by Young's formula

$$\left(\frac{\mu - 1}{\mu + 1} \right)^2.$$

So that really the minima give the gist of this part of the investigation.

For example, taking the surface e immersed in liquid No. 1, the relative index is $1.596/1.461$, or 1.0924 , and the reflexion is calculated as 1.95×10^{-3} . The observed intensity, corrected for what occurs internally, is 2.0×10^{-3} , and is therefore in good agreement.

The enhancement of reflexion in a liquid of index equal to that of the body of the silica is enormous. If we start with commercial polish it is 60 times, or if we start with a surface cleaned by hydrofluoric acid it is say 600 times or more. These conditions are of course the best for observing it. But the enhancement is so great that it should be readily visible even without a compensating liquid. The reflexion in air from a surface of index 1.596 should be 5.3% and therefore easily distinguishable from the reflexion of silica glass of index 1.46, which is only 3.5%. In fact, a silica surface, half treated and half left for comparison, shows the enhanced

reflexion very conspicuously. The back surface should of course be covered with black varnish to isolate the front reflexions which are to be compared and it is important to get the edge of the treated half as sharp as possible, to bring out the contrast. Such a specimen readily lends itself to lantern projection.

Since the modified layer is made from silica glass, and since apparently nothing else enters into its composition, it would seem that it must be a modification of silica. This is in accordance with the fact that it is unaffected by ordinary acids but removed by hydrofluoric acid. But the high index of the layer, 1.61 or more, makes it impossible to identify it with any of the recognized modifications of silica. Thus

Crystalline quartz (doubly refracting) has a mean index about 1.55,

Tridymite has index 1.48,

Crystobalite 1.43;

Opal, hyalite, etc., 1.45 at the most.

The modified surface was kindly examined for me by Professor G. I. Finch in the electron diffraction camera. No evidence was obtained of any crystalline layer.

5—CRYSTALLINE QUARTZ

This material does not show the large variations of reflexion that are found with fused silica. It cannot, however, be treated with the same vigour, since the heat of friction on a cloth or felt polishing wheel is very liable to result in cracking. Notwithstanding this limitation I was able to satisfy myself that the changes of reflecting power which can be produced are comparatively small. To annul the reflexion for green light of a quartz wedge (cut perpendicular to the optic axis) the appropriate liquid was found to be a mixture of 36 c.c. carbon disulphide with 64 c.c. benzene. The commercial polished surface (immersed) gave a dull red reflexion in white light. Using the green glass filter the reflexion was insensible with the arrangements used. The sector value was less than 0.1 mm. and the intensity less than 3.5×10^{-6} of the incident light.

No modification of the properties of the surface was obtained by hand polishing with dry rouge on a paper polisher. By rubbing on the felt wheel with "abradum" an effect was produced, but it was not conspicuous enough to shift the minimum reflexion to another of the stages represented by the series of liquids. The location of a flat minimum is not however a very sharp criterion, and it was found that *before* the minimum the treated

surface gave the larger reflexion while *after* it the original surface (i.e. the untouched face of the wedge) did so. This indicates that the surface has been modified in the direction of increasing the refraction. But the effect is small. We may contrast the sector value of 0.1 mm. (intensity 3.5×10^{-6}) for treated quartz with the sector value of 60 mm. (intensity 2.10×10^{-8}) for treated silica glass, each examined in a liquid of the same refraction as the interior.

There is however no doubt of the reality of the effect with quartz. This is conclusively proved by the colour phenomena which will be referred to later.

6—GLASS

Some experiments have been made with ordinary glass to test whether it would behave in any degree like fused silica. An ordinary crown glass spectacle prism of 2° angle was used. The liquid to annul the reflexion was made up by mixing 5 c.c. of carbon disulphide with 11 c.c. benzene. In this liquid the original surface gave a sector of 0.25 mm. (intensity 8.8×10^{-6}) which could have been got lower if the surface had been treated with hydrofluoric acid. By 30 min. treatment on the felt wheel and "abradum" the value was got up to sector 3.1 mm. (intensity 1.09×10^{-4}). Another 30 min. work did not appreciably increase this value. The effect, though not nearly so striking as with fused silica, is still marked enough. Owing to the higher initial index of glass, we should scarcely expect that the enhanced reflexion would be so conspicuous as with silica. But besides this higher initial value, the ultimate value attained is much less (1.541 as against say 1.596) so that, due to the combined causes, the effects are much less conspicuous.

I have noticed casually that some specimens of plate glass, and also an old optically polished surface of flint glass (telescope lens) had considerable reflecting power in a liquid of index equal to the interior. This could be got rid of by dipping in dilute hydrofluoric acid, or by standing for several days in a concentrated solution of caustic potash.

7—COLOUR PHENOMENA

As already mentioned, the refractive indices given are for green light, and green light is used for the photometric determination of reflexion, a green glass being held over the eye.

The colour effects with white light are in some cases very conspicuous,

and force themselves on the observer's attention. In Table III the colour effects are to some extent recorded. The light reflected from the immersed specimen is compared with the light reflected from a silica surface in air.

R and *RR* mean that it is in varying degrees redder.

B, *BB* and *BBB* mean that it is in varying degrees bluer.

N (neutral) means that there is no noticeable difference of tint.

If we examine the case of glass, we notice that the change over from red to blue occurs at the minimum of reflexion. This applies whether the surface is modified or not. After treatment on the wheel with "abradum", the minimum reflexion and the change of colour are shifted *pari passu* to a more refracting liquid.

In the case of crystal quartz essentially the same is observed. In this case the change in refraction produced by the wheel is not great, but the colour phenomena make it very definite, more so indeed than the reflexion measurements taken by themselves would do. We find that in a suitable liquid index about 1.55, the original surface gives a faint but very blue reflexion, while the modified surface gives a faint red one. The two images from the faces of the wedge are seen side by side, and it is obvious at a glance that the surfaces do in fact differ as a result of the treatment.

With the various specimens of fused silica these colour changes are less conspicuous, and they do not mark the minimum reflexion in the same way. The unmodified surface was not examined in liquids less refractive than itself. The most highly modified surface shows a marked red colour right through the minimum for a considerable range on either side.

The explanation of this change of tint at minimum in such a case as crystalline quartz is as follows. Let a curve be constructed showing as abscissae positions in the spectrum (wave lengths) and as ordinates the intensities of reflexion $\left(\frac{\mu-1}{\mu+1}\right)^2$ for a given immersion liquid, μ being the index of the liquid relative to the surface under examination for reflexion.

Then, if the liquid is adjusted to have minimum reflexion in the green, the curve will be somewhat as shown by *gGg'* (fig. 1)

If we make the liquid (mixture of bisulphide and benzole) more refracting by adding bisulphide, then, as experiment shows, the point at which it compensates quartz is shifted towards the red, and the curve will be shifted to a new position, as indicated by the curve *rRr'*. With this liquid if we observe in the red we shall get small reflexion (vanishing ordinate at *R*); in the blue, larger reflexion, indicated by the ordinate *Br*. Similarly, if we

make the liquid less refragable by adding benzole, the point of compensation will be shifted towards the *blue* as in bBb' . We get small reflexion in the blue (vanishing ordinate at B) and large in the red indicated by the ordinate Rb' .

The above explanation should be applicable to all cases when the liquid is more dispersive than the solid. The last column of Table III shows that it applies to a surface of flint glass (deprived of skin by treatment with hydrofluoric acid) which has minimum reflexion in liquid No. 12, index 1.596, which liquid also compensates the refraction through the body of the material. This approximates closely to one of the modified surfaces of silica (column e in Table III). The colour is observed to change from red to blue as we pass through the minimum with increasing refraction.

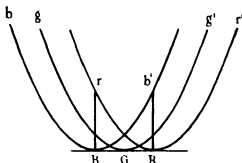


FIG. 1

It is very improbable that the modified surface of silica should be more dispersive than a liquid consisting mainly of carbon disulphide; but even if it were so, the observed behaviour would not be explained. For then with increasing refraction of the liquid, the colour should change from blue to red as we pass through the minimum. But in fact it remains red on both sides of the minimum. It is to be noticed that the persistent reflexion at minimum is larger in this case than in the cases where the expected change of tint occurs, and we may take the observation as indicating that this persistent reflexion is red, and swamps all other colour effects. On the view explained in § 4, the persistent reflexion is due to the more or less gradual transition between modified and unmodified silica at the internal surface of the modified layer. This gradual transition would favour red reflexion at the expense of blue.

8—THICKNESS OF THE MODIFIED LAYER

It is evidently important to determine this thickness. The method used is to remove the modified layer by immersion in hydrofluoric acid, and to

determine how much is removed. This may be done by weighing, but the differences of weight to be dealt with are rather small, and it was found preferable to use an interference method to measure the depression of the eroded portion below the original level.

The method was finally developed as follows. A long rectangular piece of silica was made as highly reflecting as possible by the treatment previously described. It was then coated with beeswax and transverse lines were ruled so as to divide the length alternately into strips 5 mm. and 2.5 mm. wide. The wax was removed from the latter only, leaving the edges as clean cut as possible. Commercial hydrofluoric acid, diluted 12-fold with water, was used. The acid was in a vertical lead cylinder, quite full, and the strip could be smoothly lowered into it by a holder with a rack movement. The first narrow strip was immersed and left a minute. The silica was then lowered so as to immerse the second narrow strip, and another minute given, and so on, until the last strip had been given a minute's immersion. In this way the successive strips (inverting the order) had a total immersion of 1, 2, 3, 4 min.,

On examining the reflexion in a liquid of equal index it was found that the 1 min. strip showed considerable reflexion, the 2 min. strip showed less, and the 3 min. and succeeding strips practically none. It is concluded that about 2.5 min. is required to remove the layer. To determine the depth to which these strips were eroded, the silica was clamped to a glass flat, and the interference fringes observed in white light, manipulating the clamps so as to get the fringes perpendicular to the strips. The fringes were observed to be dislocated when they crossed a strip, and since the individual fringes can be identified in white light by the purity of the colour it was possible to be certain that the amount of the dislocation was only a fraction of a period. To determine its exact amount it is better to use monochromatic light; the estimate made by measuring a photograph was about 0.1 of a period (blue mercury line) and there is no doubt of the substantial correctness of this result. It is believed, however, that an estimate got in the following way was better. We assume that the depth of erosion is proportional to the time. This assumption is no doubt open to the criticism that the modified layer may be eroded more quickly or less quickly than the interior; nevertheless I believe that it is nearly correct. A dislocation of half a period is easily identified, a black fringe on the eroded part being brought exactly half way between two black fringes on the uneroded part. This occurred after 10 min immersion. We conclude that in 2.5 min. the dislocation should be $\frac{0.5 \times 2.5}{10} = 0.125$ of a period. This is in fair agreement

with the more direct method, and is probably more accurate. Green mercury light was used in this case, and 0.125 of a period indicates a depth of 0.0625λ or 341 Å.

9—APPLICATION TO THE STUDY OF CONTACTED SURFACES

We have seen in detail that polishing processes in general tend to form a layer on the surface of silica differing in reflecting power from the interior, and to a variable extent according to the exact technique of polishing employed. It was thought desirable to show by direct experiment that this in fact conditions the variable reflexion from a contacted interface of silica.

A pair of commercially polished wedges were contacted. The reflexion gave sector opening 1 mm. (intensity 3.5×10^{-3} of incident light). The wedges were then taken apart, and one of the surfaces originally contacted was repolished for 20 min. on a paper polisher with rouge. The surfaces were then recontacted, and another test made. The sector opening was now 20 mm. (intensity 7.0×10^{-4} of incident light) and has thus increased 20-fold.

The surfaces were then taken apart, and the other one similarly treated. On recontacting, the sector value had now gone down to 7 mm. (intensity 2.45×10^{-4} of incident light).

These tests verify the point as clearly as could be wished. Treating one surface alone increases the reflexion greatly. As might be expected, treatment of the second surface reduces it again; but not to its original low value, for even if interference were complete, so that the two outer modified surfaces fused together and became optically equivalent to no interface at all, we should still have a stratum of modified material inside the unmodified mass; and this should give reflexion.

There can be little doubt that this is what actually happens when two contacted silica surfaces are "adhesed" by the application of heat. As was shown in the former investigation (Rayleigh 1936, p. 344) reflexion from the joint still subsists under these circumstances though the mechanical union appears to be complete. If this view of the matter is right, the modified surface should be able to survive heating; and as we have already seen (p. 511) it can in fact do so.

10—MEAN DISTANCE OF CONTACTED SURFACES

We have seen that the reflexion from contacted glass or fused silica surfaces is variable, on account of the skin differing in optical quality from

the interior. It is necessary to be certain that there is no such skin before anything can be inferred from the reflexion of the contacted interface as to the true distance between the surfaces. As already explained, the test is that there should be no reflexion when the surface is immersed in a liquid of the same index as the interior

This result can be secured by dipping for a short period in hydrofluoric acid, and my hope was that by contacting surfaces so treated a good experiment could be made. At a first glance the etched surfaces do not seem to have lost much of their polish, and no difficulty is found in contacting them, if they possess the necessary original flatness. But the contacted surface is full of defects, and scatters so freely that the specular image is surrounded by a flare of diffused light.

This behaviour of glasses treated with hydrofluoric acid might to some extent be anticipated from observations made many years ago by my Father (Rayleigh 1893, p. 59 of the reprint in *Scientific Papers*). He found that the acid developed scratches and pits on the surface which were not initially visible, and which were presumably the same as those which had seemed to have been obliterated in the polishing process. They had been as it were skinned over, but the acid removed the skin (see also Beilby 1921, p. 110).

An attempt was made to select a small part of the surface which might be free from defects of this kind. The glasses used were parallel, not wedged, and in order to separate the interface reflexion from the front and back reflexion, a narrow beam only was used, an illuminated slit being focussed on the glass surface by a lens of 1 metre focal length and 1 cm. aperture, and the reflexion being observed at about 13° from the normal. The interface reflexion could then be seen separate from the reflexion from the two outer surfaces and between them; but it was necessary to mitigate the brightness of these latter, which produced far too much glare. The back surface was coated with black cellulose lacquer, which made the brightness comparable with that of the interface, and the front surface had a block of glass 12 mm thick cemented to it with canada balsam

We have then in order:

Reflexion from black lacquered back surface (a) faint.

Reflexion from contacted interface (b) faint.

Reflexion from balsam (c) faint.

Reflexion from front surface (d) bright.

They are spaced as shown in fig. 4, the reflexion from (d) being much farther off than the others. It could readily be screened off, and did no harm.

The reflexion from (*b*) could be observed with a magnifier of 5 cm. focal length, focussed on the interface. It also served as a field lens, as shown in fig. 4, which indicates the course of the rays par-axial to the lens which are reflected from the various surfaces. Suitable arrangements allowed the glass to be moved parallel to itself, so that every part could be examined separately. But each part was found to be covered with scratches and pits which appeared like bright stars in the field, and made any satisfactory examination of the background impossible. The same applied to glasses deprived of their skin by prolonged soaking (20 h.) in concentrated caustic potash solution. [On the other hand, when glasses which had not been treated with hydrofluoric acid were examined in this way, the interface

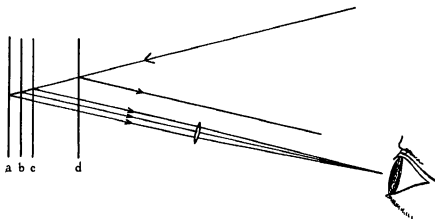


FIG. 4

reflexion (*b*) could be examined very satisfactorily, and could be compared with that from the back surface (*a*), which served as a kind of standard. A neutral glass of known opacity could be introduced, so as to equalize the intensities]

I was therefore compelled to give up chemical methods of removing the skin, and was for some time at a loss how to proceed.

However, it was recalled that crystal quartz surfaces commercially polished gave a practically insensible reflexion in the liquid, at that part of the spectrum (green) for which the refraction was equalized to that of the interior.

Two crystal quartz wedges with pretty accurately flattened surfaces were tested in liquid, and found to give the ordinary dull red reflexion, which viewed through green glass, had a value not more than 10^{-4} of the

reflexion of silica glass in air or 3.5×10^{-6} of the incident light. This was true of each of the surfaces. The amount of reflexion for green was too small to be observed very definitely, but did not exceed the amount stated.

These prisms were contacted, which was found to be definitely easier with crystal quartz than with glass. No benzene was used, the cleaned surfaces being simply slid together. The interface reflexion was measured in the usual way without difficulty. The intensity compared with the reflexion of silica glass in air was found to be 1.2×10^{-3} , or, compared with the incident light, 4.2×10^{-6} . Three pairs of surfaces were tried, always with about the same result. This is quite ten times as bright as the reflexion in a liquid of equal index, and it would seem that skin effect cannot be important in this case, and that the reflexion must be genuinely attributable to the mean distances between the contacted surfaces differing appreciably from zero. The distance is given by

$$t = \left(\frac{I}{e^2} \right)^{\frac{1}{2}} \frac{(1 - 2e^2)^{\frac{1}{2}}}{2K},$$

where I is the reflected intensity, e the factor by which the amplitude of the incident wave is altered on reflexion, and $K = \frac{2\pi}{\lambda}$.

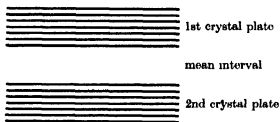


FIG. 5

In the present case $e^2 = 4.53 \times 10^{-3}$, $K = 1.15 \times 10^5$, and $I = 4.2 \times 10^{-6}$. Substituting these values, we get

$$t = 1.26 \times 10^{-7} \text{ cm.} = 12.6 \text{ \AA.}$$

Now this is a considerable multiple of the distance apart of the atomic layers in the quartz crystals themselves. The plane layers perpendicular to the axis of the crystal containing the silicon atoms are $\frac{5.375}{3} = 1.8 \text{ \AA}$ apart (Bragg and Gibbs 1925, p. 405). The interval between the surfaces as judged by the optical test is 7 times this amount, and is shown to scale in fig. 5.

We have to consider why the contact is not closer. It can hardly be supposed that the molecular forces are effective over so large a range (Rayleigh 1936, p. 347), and yet we have a strong attractive force between the surfaces. It would seem that the answer must lie in this, that full contact is at a few places only, the artificial polished surface not being molecularly flat like an ideal cleavage face, but having irregularities of a good many times the atomic spacing; so that intervals of this order will occur if it is placed against a true plane, or if as in the present case against a similar surface.

As regards the structure near the surface Beilby (1921, p. 114) considered that crystalline quartz flowed in an amorphous form in the polishing process, but recent observations of G. I. Finch (1936, p. 1010) by electron diffraction show clearly that the polish layer in quartz is crystalline, and, if I understand rightly, it must be in crystalline continuity with the interior.

The observations of the present paper are in harmony with this. It appears certain that if an amorphous layer having the properties of fused silica were present, it could not give the low reflexion observed when a polished crystal quartz surface is examined in a liquid of index equal to that of the interior of the crystal, for the reflexion of a fused silica surface in this liquid is 150 times greater. Of course, the possibility of an amorphous layer of a different character, having the index of crystalline quartz, is not excluded by this kind of evidence. But such a layer, if it existed, would not vitiate the optical determination of the mean distance apart of the contacted surfaces.

It is true that a slight increase of reflexion is observed in the polished surface (immersed), but this is a small residual effect, and may perhaps be attributed to mechanical stress or some other secondary cause.

11—MISCELLANEOUS NOTES

It was shown (Rayleigh 1936, p. 335) that the work done per square centimetre in stripping two contacted glass surfaces is 71.2 ergs. We may regard the stripping of two contacted surfaces as the creation of a new free surface. The work done per square centimetre of the new surface will be half this amount, i.e. 35.6 ergs. The process is analogous to the creation of a liquid surface which requires an amount of work measured by surface tension \times length. Thus the creation of a new water surface requires 73 ergs per square centimetre and of a new alcohol surface 22 ergs per square centimetre. These are of the same order as the value found for glass. Too

much stress must not be placed on this, in view of the probability above arrived at that the surface is patchy, the closest contact being local only.

It is worth noting that while quartz crystal and vitreous silica are more easily contacted than glass, yet there does not seem to be any exceptional affinity of a glass surface for another similar to itself; and in fact if we start with a glass plate it is easier to contact a piece of silica glass on to it than to apply a piece of glass similar to the first piece.

In carrying out the above work I have received valuable assistance from Mr. R. Thompson.

12--SUMMARY

The reflecting power of a polished silica surface examined in a liquid having the same refractive index as the body of the material varies widely according to the treatment.

Surfaces polished by a process which removes material rapidly, or surfaces washed in dilute hydrofluoric acid, do not reflect appreciably in the liquid.

Surfaces polished by methods which do not quickly remove the material may reflect (in the liquid) as much as 0.28 % of the incident light. A kind of burnishing seems to take place which modifies the surface, and may bring up its refractive index from 1.461, the ordinary value, to as much as 1.6, quite as high as light flint glass, and much higher than any known variety of silica. These effects are found in a less degree in ordinary glass, and in a very much less degree in crystal quartz. In normal cases the reflected light changes in tint from red to blue as the refractive index of the immersion fluid is increased through the critical value for minimum reflexion. The modified silica surface is anomalous in this respect, the reflexion being red on either side of the minimum value. An explanation is suggested for this.

The thickness of the modified layer was measured as 0.06λ , when λ is the wave-length of green light in air.

The variable reflecting power formerly found from the interface of two silica or glass surfaces in optical contact is now explained. It can be controlled by using the present results. Contacted surfaces of crystal quartz give a reflexion which is practically independent of the way in which the surfaces have been polished. The mean distance between the two crystals is found to be about 7 times the spacing of the layers of silicon atoms in the crystals.

REFERENCES

- Beilby, Sir G. T.* 1921 "Aggregation and Flow of Solids." London: Macmillan and Co.
Bragg, Sir W. H and Gibbs, R. E. 1925 *Proc. Roy. Soc. A*, **109**, 405-27.
Finch, G. I. 1936 *Nature, Lond.*, **138**, 1010.
Rayleigh (Lord, the late) 1893 *Proc Roy Instn.*, **14**, 72-8. Also *Scientific Papers*, **4**, 54-9.
Rayleigh, Lord 1936 *Proc. Roy. Soc. A*, **156**, 328-49.

The Photo-Electric Measurement of the Diurnal Variations in Daylight in Temperate and Tropical Regions

BY W. R. G. ATKINS, Sc.D., F.R.S. (*Plymouth*), N. G. BALL, Sc.D. (*Colombo*), AND H. H. POOLE, Sc.D. (*Dublin*)

(Received 19 March 1937)

In previous papers we have considered the standardization of photo-electric cells for the measurement of daylight (Poole and Atkins 1935), the daily and seasonal changes (Atkins and Poole 1936*a*) in daylight in England, and the relation of these photo-electric scales to the visual scale of photometry (Atkins and Poole 1936*b*).

Though measurements of total radiation had been made in the tropics and though certain spectrophotographic records had been obtained, no measurements of daylight had, as far as we knew, been made by photo-electric methods strictly comparable with those used in Europe and with the newer types of cell. Cells were accordingly standardized against those used at Plymouth, and were taken abroad for use.

THE PHOTO-ELECTRIC CELLS

These were of two types: Burt vacuum sodium cells, and, later on, Weston selenium cells of the rectifier type.

(a) *The Sodium Cell*

Evidence has been adduced for the constancy of the sodium cell B 299, which has been exposed almost continuously since 1920 on the laboratory

roof at Plymouth attached to a Cambridge thread recorder. A similar cell, B 224, which had been used on and off for several years, was restandardized by comparison with B 299 in daylight before and after use in Ceylon, and one may conclude that B 224 went through its exposure to full tropical sunlight quite unaltered. It was shown (Poole and Atkins 1935) that B 299 had a sensitivity of 3.8 kilolux/ μ A on 1 May 1934, for diffuse bright mixed daylight, on the carbon arc potassium cell scale, which is very close to the corresponding sodium cell scale. Taking the ratio of B 224/B 299 as 2.70, the sensitivity of the former is accordingly 1.41 kilolux/ μ A for diffuse light or 1.33 for vertical illumination.

Evidence previously adduced led us to conclude that these cells exhibit a rectilinear relation between current and intensity.

Though, for an equal energy spectrum, the sodium cell has its maximum sensitivity at about 360 $m\mu$ on a rather flat curve, we have shown (Atkins and Poole 1936*b*) that for the spectral energy distribution of mean noon sunlight it has a sharp maximum at 410 $m\mu$. It is accordingly registering chiefly a spectral region where the eye is of very low sensitivity, and where atmospheric impurities are likely to exert a relatively great effect.

Owing to the feeble intensity in the region of the ozone absorption band, as compared with the longer wave-lengths to which the sodium cell is sensitive, variations in the ozone content will not cause any appreciable alterations in the readings of the Burt cell.

(b) *The Selenium Rectifier Cell*

The advent of the selenium rectifier cell provided an instrument giving a photometric scale in tolerably good agreement with the visual when used for mixed daylight (Atkins and Poole 1936*b*). For this reason, and for the purpose of comparison with the results obtained in the violet, such a cell Weston 22302/7 was sent to Ceylon after comparison with our own W 21104/2, for which the curvature of the illumination/current relation had been determined (Poole and Atkins 1935). The sensitivity and curvature of these cells were nearly the same. That their colour sensitivity also was much the same is indicated by the fact that the transmission of the same 2 mm. R.G. 1 (red) Jena filter was 18.22 and 18.18% for these cells, respectively, in mixed daylight, which of course varies somewhat in colour; another pair of determinations gave 17.60 and 17.42%. Since the cells were compared in winter daylight, doubts were felt as to whether the curvature correction for W 22302/7 would hold at the high intensities to be expected in Ceylon, and beyond the range over which W 21104/2 had been studied. An additional opal-flashed (opal) glass disk was

accordingly sent to Ceylon. It is obvious that any change in the curvature would lead to alterations in the apparent transmission of this additional diffusing glass. Very careful determinations of its transmission were made in Ceylon; using the mean transmission of the opal, as obtained for the lower illuminations, to calculate the higher, the results obtained only differed from those found by extrapolation of the curvature formula by $+1.30$ to -2.16% , with average -0.70 . Thus the values obtained using the curvature formula for the highest illuminations average less than 1% too high. The error is small, far smaller than uncertainties in the reflexion loss factor, and has been neglected. These determinations afford good evidence that even in tropical sunlight temperature effects are not serious, using a 10-ohm microammeter to measure the current; with high-resistance circuits such errors are serious (Poole and Atkins 1933*a*, fig. 3).

The 10-ohm instrument, with a special shunt (Poole and Atkins 1933*a*, fig. 1) giving 1:1, 5:1, 10:1 and 50:1 ratios, while keeping the total resistance at 10 ohms throughout, was used with the selenium cells.

The galvanometer for use in Ceylon was standardized at 24°C ., but the instrument used here was tested at 17°C . In the open a temperature error may have been introduced in the galvanometer readings, but the concordance of the results indicates that it was unimportant.

There is, however, a fatigue effect, almost absent from some cells, but shown by others even in England. When the cell is shaded from the sun and then exposed the initial reading is somewhat higher than that attained after 10–20 sec., this decrease amounted to about 1.5% at noon in Ceylon, and readings given were those for the steady state. It is shown also with a cell continuously exposed to sunlight and newly connected to the galvanometer, which appears to indicate that it is a polarization rather than a temperature effect.

(c) *Comparison of Sodium and Selenium Cells*

For mixed daylight from sun and blue sky with some white clouds at Plymouth the selenium scale is often very close to our old potassium cell-carbon arc scale which is the basis of the sodium cell readings. On the other hand, under certain conditions the relative proportions of "violet" (including blue, and near ultra-violet), to which the sensitivity of the sodium cell is confined, and "yellow" (including red and green), to which the selenium cell is chiefly sensitive, vary sufficiently to cause wide differences in the illuminations recorded by the two cells.

It is not always easy to predict *a priori* the relative intensities of "violet" and "yellow" in daylight. Thus, we obtained almost identical ratios for

the readings of sodium and selenium cells at Plymouth on 22 February 1934, with low-angle sun (about 20°) and clear sky, and on 1 May 1934, with high-angle sun, blue sky and white clouds. With high-angle sun and clear sky on 11 June 1936, at Plymouth, the reading of a sodium cell was relatively 10% low, whereas under apparently rather similar conditions in Dublin on 4 June 1934, the ratio of the readings of sodium and selenium cells rose from unity at 11 a.m. to 1.25 at 1 p.m. and to 1.4 about 3.15 p.m. at which time, with the sun at about 45° altitude, the sodium cell recorded over 120 kilolux. We have also obtained some very high readings with the sodium cell and recorder at Plymouth, on days with clear sky and white clouds. Simultaneous readings of the selenium cell are not generally available.

As previously mentioned, comparison of B 224 with B 299 gives the sensitivity of the former as 1.41 kilolux/ μ A for diffuse light. On comparing B 224 with a selenium cell (W 21104/2), the value 1.44 kilolux/ μ A was found at Plymouth in mixed diffuse daylight on 28 December 1933. When compared in Colombo with another selenium cell (W 22302/7, previously standardized in mixed daylight against W 21104/2), on 6, 8 and 10 April 1935, with a clear blue sky and sun at $89-90^\circ$ altitude, B 224 gave a constant of 1.77 kilolux/ μ A for what was almost vertical sunlight, the sky effect being relatively unimportant. Thus the sodium cell is relatively less sensitive to sunlight.

THE VERTICAL ILLUMINATION IN THE VIOLET FROM SUN AND SKY FOR DIFFERENT SOLAR ALTITUDES

Observations were made with the sodium cell, at Colombo and elsewhere in Ceylon, and in fig. 1 these have been plotted against the sun's altitude. No correction has been introduced for the fact that the earth is in aphelion early in July and in perihelion early in January; this variation causes a 7% increase in solar radiation in January as compared with July.

It may be observed that for any altitude a number of values are possible, depending upon the clearness of the air and the reflexion from clouds. The possibility of variations in the sun's emission must not be left out of consideration. According to Abbot (1929) the range of the sun-spot effect on solar radiation is approximately 2%. In addition, day to day observations show fluctuations of 1-3%. Further, Pettit (1932), on Mt. Wilson, has observed large percentage solar changes in the extreme ultra-violet spectrum; the ratio of wave-lengths $0.32-0.5\mu$ varied from 0.95 to 1.56. These changes show positive correlation with the variations in total solar

radiation. It may be imagined that when specially clear atmospheric conditions coincide with a maximum in the sun's short wave-length emission, the variation produced is very considerable.

The full-line curve in fig. 1 is taken from fig. 5 and represents similar observations with the selenium cell. It is obvious that the values obtained with the sodium cell are the lower. The ten dots shown in fig. 1 represent observations at Plymouth on 11 June 1936. These lie well among the

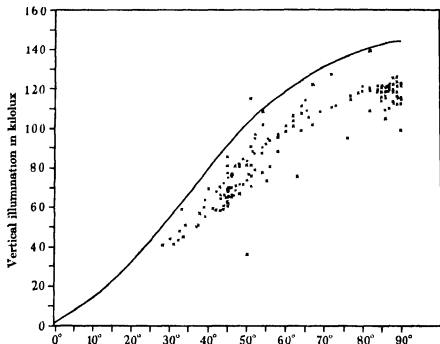


FIG. 1.—Vertical illumination from sun and sky in kilolux, as obtained using a Burt sodium cell, plotted against the altitude of the sun. The crosses show results obtained in February, March, April, September and November 1935, also in March and April 1936. The ten dots show results obtained with a similar cell at Plymouth on 11 June 1936. The full-line curve has been reproduced from fig. 5 and represents similar measurements with the selenium cell.

Colombo observations, and were made with B 209, against which the Colombo cell B 224 had been, and was subsequently, standardized. The maximum observed at Colombo, 139 kilolux, is due to the presence of clouds, and is well above the usual value with the sun at 90°, i.e. about 120 kilolux. Table III of Atkins and Poole (1936*a*) shows that the Colombo maximum recorded was exceeded from April to August 1930 at Plymouth, and was even exceeded by the mean daily maximum in April, May and July. The recorder observations, however, preserve records of unusually

bright moments which are likely to be missed in a less extensive series of observations.

Table I shows the values for sun, V_s , sky, V_d , and their ratio, also for $V_s + V_d$ for 45° altitude at Colombo. It may be seen that $V_s + V_d$ is no more constant than at Plymouth, though its absolute magnitude is much less. The Plymouth values for 1934 are obviously more truly comparable with those obtained in the same year in Colombo, as they are more nearly simultaneous and the sodium cells were compared before and after the two sets of measurements. Even these 1934 Plymouth illuminations are well above those obtained in Colombo, though much lower than the 1930 results (see Atkins and Poole 1936a, Table VII). Our results showing a relatively lesser intensity in the violet end of the spectrum in the tropics are quite in accord with the numerous spectrophotographic records given by Dorno (1911, 1930).

TABLE I—OBSERVATIONS AT COLOMBO, CEYLON, WITH SODIUM CELL
VERTICAL ILLUMINATION IN KILOLUX WITH SUN AT 45° , FROM SUN,
 V_s , FROM SKY, V_d , SUN/SKY RATIO AND $V_s + V_d$, AURÉN'S E_s UNIT
PLYMOUTH AND DUBLIN RESULTS BELOW

Date	V_s	V_d	V_s/V_d	$V_s + V_d$	Conditions
20. ii. 1934	—	—	—	81.3	Blue sky, some low white clouds
29. iii. 1934	43.6	19.5	2.23	63.1	Blue sky, some scattered white clouds
3. iv. 1934	39.6	22.7	1.74	62.3	Blue sky, very faint haze
14. xi. 1934	—	—	—	65.5	Clear blue sky
16. xi. 1934	—	—	—	60.6	Blue sky, scattered white clouds
16. xi. 1934	—	—	—	69.7	Blue sky, a few scattered white clouds
17. xi. 1934	—	—	—	63.1	Blue sky, practically cloudless
19. xi. 1935	—	—	—	70.2	Blue sky, large scattered clouds
18. iii. 1935	40.5	22.9	1.77	63.0	Blue sky, scattered wisps of cloud
18. iii. 1935	42.8	35.0	1.22	77.8	Blue sky, cloud in east and scattered clouds
19. iii. 1935	43.6	17.6	2.47	60.8	Blue sky, white cloud near east horizon
19. iii. 1935	50.0	35.8	1.39	85.8	Sun clear, but numerous white clouds
20. iii. 1935	46.1	20.4	2.26	66.0	Blue sky, some white clouds
19. iv. 1935	39.3	29.7	1.32	69.0	Sun clear, large white clouds near it
Mean	43.2	25.4	1.70	68.45	
1930	—	—	—	100.4	Plymouth, mean of 32 obs., max. 137.9, min. 62.7
1934	—	—	—	85.5	Plymouth, mean of 5 obs. with nearly clear sky, in April, max. 96.5, min. 76
1934	—	—	—	83.1	Plymouth, mean of 20 obs. with nearly clear sky, in May, max. 98, min. 69
4. vi. 1934	76.1	44.9	1.69	121.0	Dublin, $\alpha = 45^\circ 30'$, clear blue sky, wind east
11. vi. 1936	—	—	—	75.5	Plymouth, very clear blue sky

THE VERTICAL AND TOTAL ILLUMINATION FROM SUN AND SKY AS
DETERMINED WITH THE SELENIUM CELL AND GLOBE PHOTOMETER

In our previous paper (Atkins and Poole 1936*a*, figs. 5 and 6) measurements of V_d and V_s , and their sum, were compared with those for the total illumination I , and its components I_d and I_s , values of I being obtained by means of the globe photometer the readings of which were shown to be independent of the altitude and azimuth of the source. These observations at Plymouth (50° 22' N. and 4° 08' W.) only extended up to a solar altitude of 22°. Fig. 2 shows further observations up to 63°; those for sun alone have been omitted for clearness, as they are the differences of I and I_d , and of V and V_d , respectively.

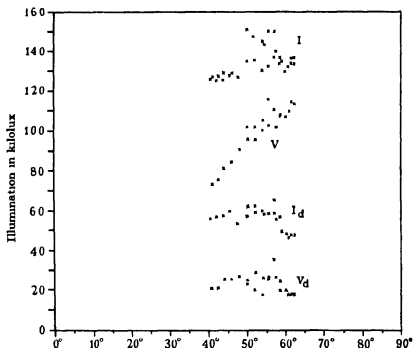


FIG. 2.—Total and vertical illumination from sun and sky, I and V , and from sky alone, I_d and V_d , as obtained with selenium cells and globe photometer, at Plymouth, on laboratory roof, 24 June 1936. Sun clear, clear blue sky overhead, white clouds around, haze over sea, wind south-south-west, very light.

For a uniformly illuminated sky and no reflexion from the ground or sea $I_d = 2V_d$. Actually values ranging from $1.86V_d$, when a white cloud raised both, to 2.71 were obtained, including more typical values 2.27 , 2.48 and 2.66 . The effect of the high-altitude illumination from bright clouds tends

to exalt V_d , and the reflexion from the ground and from the sea tends to raise I_d , the latter effect preponderating as a rule. Moreover, a clear blue sky is considerably brighter near the horizon than near the zenith in azimuths remote from that of the sun (Poole and Atkins 1933*b*).

The most striking result is that whereas alteration in solar altitude naturally has a great effect upon V , the corresponding values for I alter but little. Here the purely geometrical effect is eliminated, variation in altitude is thus effective through an alteration in the length of the path of the sun's rays through the atmosphere, which becomes more important at low solar altitudes. The biologically important result is that an approximately spherical tree or bush may receive a total illumination which is approximately constant during the middle of the day, whereas the illumination on the sea varies considerably over the same period.

The maximum value of I obtained on a clear June day with white clouds was 151 kilolux, with the globe photometer, with simultaneously 116 kilolux for V , a value far surpassed on other occasions.

THE VERTICAL ILLUMINATION AT DIFFERENT SOLAR ALTITUDES AS DETERMINED WITH THE SELENIUM CELL

Fig. 3 shows the illumination at sunrise on the east coast of Ceylon. As reckoned from the time when it was too dark to take measurements, up to the appearance of the sun's rim, "twilight" lasted for 16 min. The illumination as the rim appeared was 702 lux, which may be compared with 696 lux found for the moment the sun's rim disappeared at Plymouth (Atkins and Poole 1936*b*). According to fig. 3 (Atkins and Poole 1936*a*), "twilight" at Plymouth lasts about 40 min. in February and October.

Fig. 4 gives the illumination at Plymouth, up to 63°, the maximum solar altitude obtainable. Four Dublin observations are also shown. The effects of slightly hazy conditions and the typical east wind haze are shown by a number of low results. High values, as before, are attributable to specially clear air and white clouds. The maximum value, 155 kilolux, might be reduced to 151 kilolux by adopting a different reflexion loss factor, but as the angular distribution of the light was unascertainable with exactitude on account of the presence of numerous clouds the correction is doubtful, and is, in any case, unimportant. It may be seen that reflexion from clouds can raise the illumination above that given by the sun at 90° in Ceylon, for the full curve in fig. 4 has been reproduced from fig. 5, though it is a tolerably good representation of the results of fig. 4 itself.

Fig. 5 shows results for April and September 1935, the times at which the sun is vertically overhead at Colombo (latitude $6^{\circ} 54' N.$, longitude $79^{\circ} 56' E.$). The observations are quite concordant and in general the more settled weather conditions result in the observations for the same altitude being closer together than in fig. 4. There is no indication of any loss of sensi-

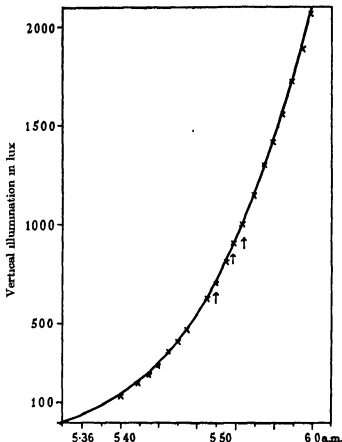


FIG. 3.—Variation of the vertical illumination, as found using a selenium cell under opal, before and after sunrise on a clear morning at Arugam Bay, Ceylon, 1 May 1935. The first arrow indicates the appearance of the sun; the second when the full disk was visible; the third when sun was clear of light cloud.

tivity in the cell. The value for 90° is about 145 kilolux as against 120 kilolux with the sodium cell in Colombo. The maximum value observed, 155 kilolux, is within experimental limits of error identical with that found at Plymouth. The outstanding result of the comparison is that with the selenium cell the general agreement between the observations in England and in Ceylon is far better than for the short wave comparison with the

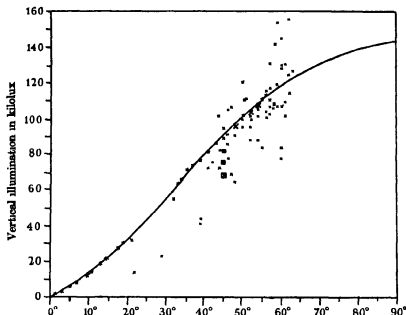


FIG. 4—Vertical illumination at Plymouth, as determined with a selenium cell during February 1934, and June and July 1936. The large crosses denote coincident observations, the crosses in squares observations at Dublin in June 1934. The full line curve has been reproduced from fig. 5.

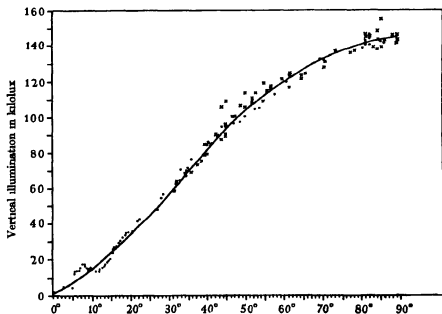


FIG. 5—Vertical illumination, as found using a selenium cell under opal, at Colombo, Ceylon, for different altitudes of the sun. The sun was clear, with a few clouds at times. The observations of April 1935 are shown by crosses, those of September 1935 by dots. The irregularity in the latter below 40 kilolux was occasioned by clouds near the horizon.

sodium cell. Many people will be surprised to notice that there are times in England when the brightness of the light is equal to or exceeds that of Ceylon with the sun in the zenith. The very high values, 250–325 kilolux over 6-hour periods, reported by Teegan and Rendall (1930*a, b*) for Rangoon were obtained on a scale widely divergent from the visual; they used a blue sensitive potassium cell, standardized against a "Pointolite" electric lamp (*viz.* a predominantly red source), to measure daylight.

Even with the selenium cell atmospheric conditions militate against more than a very approximate value being assigned to the average illumination from sun and sky together, namely Aurén's " E_s unit". Table II gives such values for different places. Even with the sun apparently clear, the value for sunlight may be doubled on a very clear day; and on the latter the sky light may be halved. These remarks may be considered as an addition to the discussion in our 1936*a* paper, § VI.

TABLE II—VERTICAL ILLUMINATION IN KILOLUX WITH SUN AT 45° , FROM SUN, V_s , FROM SKY, V_d , SUN/SKY RATIO AND $V_s + V_d$, AURÉN'S E_s UNIT, AS FOUND WITH SELENIUM CELL. OBSERVATIONS WITHIN 40 M. OF SEA-LEVEL, SAVE AT DIYATALAWA 1210 M.

Date	V_s	V_d	V_s/V_d	$V_s + V_d$	Place and conditions
1. vi. 1934	—	—	—	68.3	Dublin. Hazy blue sky, wind east, wispy clouds, horizon grey
2. vi. 1934	33.8	34.7	0.97	68.5	Dublin. Sun through high light haze
4. vi. 1934	64.9	17.3	3.75	82.2	Dublin. Clear sky, high wisps of cirrus
11. vi. 1934	54.9	20.9	2.63	75.8	Dublin. Blue sky, very high white wisps of cirrus
6. iv. 1935	79.9	15.8	5.06	95.7	Colombo. Sun clear, some hazy white clouds
8. iv. 1935	72.4	17.3	4.18	89.7	Colombo. Sun clear, some low white clouds
26. iv. 1935	73.0	16.5	4.42	89.5	Diyatalawa. Sun clear, clear blue sky
11. vi. 1936	—	—	—	85.7	Plymouth. Sun clear, very clear blue sky
18. vi. 1936	36.3	36.4	1.00	72.7	Plymouth. Sun clear, blue sky, white broken clouds
24. vi. 1936	58.5	25.2	2.32	83.7	Plymouth. Clear sun, clear blue sky, white clouds

THE RATIO OF SUNLIGHT TO SKY LIGHT, V_s/V_d , FOR VARIOUS SOLAR ALTITUDES

Fig. 6 shows observations with both sodium and selenium cells as explained in the legend. For the former the ratio never exceeds four, for the latter eight may be passed. Obviously the sky light affects the sodium

cell far more than the selenium cell. Even in England the selenium cell may show a ratio of four to six, and at Diyatalawa the very high value 6.6 was found for $\alpha = 47^\circ$. At that altitude, 1210 m., the sky was very clear. Tables I and II emphasize how variable this ratio may be even with clear sun at 45° . It would appear to be desirable to make such determinations with colour filters to limit the observations to narrow spectral bands. The

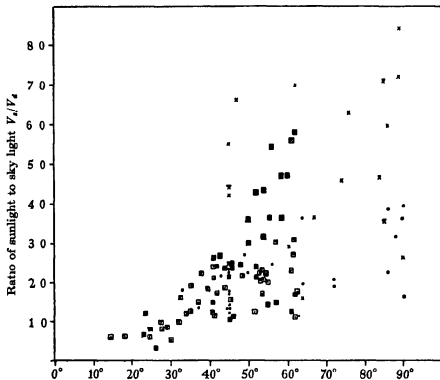


FIG. 6—Ratio of the vertical illumination from the sun to that from the sky, V_s/V_∞ , is plotted against the altitude of the sun. Values obtained in April 1935, and June and July 1936, at Colombo, Ceylon, are shown by crosses; crosses with a line at the top denote observations at Diyatalawa, Ceylon. All were obtained using a selenium cell. Dots show similar observations made with a sodium cell at Colombo, in February, March and April 1934, and in March 1935. Dots inside a square show observations at Plymouth during 1930 with a sodium cell and crosses inside a square relate to Plymouth observations with a selenium cell in June 1936.

differences between a measurement around $410\text{ m}\mu$ and one in the middle of the visible spectrum are very clearly shown by the observations already made.

We desire to express our thanks to the Government Grant Committee of

the Royal Society for defraying the cost of the special instruments used in this work. We are also indebted to Sir George Simpson, K.C.B., F.R.S., for helpful criticism.

SUMMARY

Sodium and selenium photo-electric cells were compared, and their stability tested, after which one of each type was used to measure the vertical component of daylight in England and in Ceylon.

For the same altitude of the sun the intensity around $410\text{ m}\mu$, as measured by the sodium cell, is usually less in Ceylon than in Plymouth. For 90° altitude an illumination of about 120 kilolux on the potassium cell-carbon arc scale is obtained in Ceylon. This scale agrees well in England, in bright mixed daylight, with the selenium cell artificial mean noon sunlight scale. With the selenium cell the agreement between determinations in England and in Ceylon is far closer than with the sodium cell. The value found for 90° altitude is about 145 kilolux but the maximum observation was 155 kilolux in both Plymouth and Colombo.

No evidence could be found for considering the average value of the illumination from sun and sky as more than a very rough approach to constancy with a solar altitude 45° , the individual values for sun and sky might be doubled or halved for the same altitude.

Twilight, as reckoned down to an illumination of one lux, lasts about 40 min. in Plymouth in early October and 16 min. in Ceylon in early May, but the illumination at sunrise or sunset on a clear day is the same, about 700 lux on the selenium cell scale.

The total illumination from the sky as received upon a small globe is usually more than the theoretical twice the vertical uniformly diffuse light, since reflexion from the ground or sea increases it and a blue sky is generally brighter near the horizon, so the ratio is increased up to about 2.7.

The ratio of sunlight to sky light (V_s/V_d) may be as much as 4 for $410\text{ m}\mu$ and 90° altitude, but may be over 8 for the middle of the visible spectrum.

REFERENCES

- Abbot, C. G. 1929 "The Sun", revised ed. New York: D. Appleton & Co.
Atkins, W. R. G. and Poole, H. H. 1936a *Philos. Trans. A*, **235**, 245-72.
— 1936b *Proc. Roy. Soc. B*, **121**, 1-17.
Dorno, C. 1911 "Physik der Sonnen- und Himmels-strahlung." Braunschweig: Vieweg.
— 1930 *Naturwissenschaften*, **18**, 249-52.

- Pettit, E. 1932 *Astrophys. J.* **75**, 185-221.
Poole, H. H. and Atkins, W. R. G. 1933a *Sci. Proc. R. Dublin Soc.* **20**, 537-46
— — 1933b *Sci. Proc. R. Dublin Soc.* **21**, 1-8.
— — 1935 *Philos. Trans. A*, **235**, 1-27.
Teegan, J. A. C. and Rendall, G. R. 1930a *Nature, Lond.*, **125**, 447.
— — 1930b *Indian J. Phys.* **4**, 585-8.
-

The Absorption Spectra and Photochemistry of Polyatomic Molecules Containing Alkyl Radicals

V—Vibration Frequencies and Structure

BY H. W. THOMPSON, *Fellow of St John's College, Oxford*
AND J. W. LINNETT, *Formerly Scholar of St John's College*

(Communicated by C. N. Hinshelwood, F.R.S.—Received 19 December 1936)

1—INTRODUCTION

One of the most important properties of a molecule is its ability to vibrate. A complete knowledge of the normal vibrations of a polyatomic molecule not only provides valuable indications of its structure, but also may eventually be of value in elucidating the mechanism of reactions in which the molecule takes part, for it is largely among such vibrational degrees of freedom that energies of activation may be stored.

The vibration frequencies of molecules are usually determined by spectroscopic methods, and considerations based on known Raman, infra-red, or ultra-violet spectral data often make it possible not only to determine their magnitudes but also to assign observed vibration frequencies to the specific vibrational types. It is, however, also possible, provided some knowledge of the force field existing in a molecule is assumed, to calculate the normal vibration frequencies of molecules, and at the same time to deduce the directions of the atomic amplitudes. Such calculations (Bjerrum 1914; Dennison 1926; Yates 1931; Radakowicz 1930, 1932; Lechner 1932; Howard and Wilson 1934; Sutherland 1936; Howard 1935; Sutherland and Dennison 1935) often serve to test previous empirical assignments, and

may also help in deciding or confirming the molecular structures. If the vibration frequencies of a molecule are known from experimental data, both in regard to magnitude and assignment, we can, conversely, calculate the properties of the force field existing in the molecule, this being customarily expressed by the magnitudes of the various "force constants". We might then, *a priori* at least, expect to be able to use these deduced force constants in calculating the vibration frequencies of more complex molecules of the same type and involving similar linkages. An example of this procedure is discussed in the present paper, and the extent of its limitations is revealed.

We have recently measured the ultra-violet absorption spectra of a series of polyatomic molecules in which ethyl and methyl radicals are attached to a central polyvalent atom, and have attempted to analyse some of the band systems observed (Thompson and Linnett 1934, 1935, 1936, 1937). We are here primarily concerned with the methyl and ethyl derivatives of zinc and mercury. Both zinc dimethyl and mercury dimethyl can

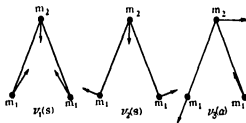


FIG. 1

be regarded to a first approximation as triatomic systems, each having three normal vibration frequencies (fig. 1). The Raman frequencies of these molecules have been measured by Gopala Pai (1935), who, also regarding them as triatomic systems, assigned three relatively intense frequencies to the three normal vibrations. The frequencies which we have previously deduced from the ultra-violet spectra agree roughly in magnitude with those measured by Gopala Pai, but the assignment inferred by us from intensity considerations in accordance with the rules deduced by Herzberg and Teller (1933) by an extension of the Franck-Condon Principle is not the same as suggested by Gopala Pai. There were already several independent reasons for preferring our assignment, but the arguments of the present paper make it even more probable.

Having established the correct assignment, the force constants of the linkages in the metallic methyl compounds can be determined and the frequencies of the ethyl compounds then calculated and compared with those found in the Raman effect, both as regards intensity and magnitude.

2—THE VIBRATION FREQUENCIES OF TRIATOMIC SYSTEMS *ABA*

Two methods may be followed in calculating the normal vibration frequencies of molecules.

The first is particularly suitable for linear molecules, and has been previously used for such cases. It makes use of the fact that in a normal vibration all the mass points are subject to the condition $\nu = \frac{1}{2\pi}\sqrt{\mu}$, i.e. all have the same "restoring acceleration". The restoring force acting on any mass point m , displaced a distance x from its rest position, is then $-\mu mx$. This quantity, taken for each mass point in turn, may be set equal to the algebraic sum of the restoring forces acting upon the particular mass arising from the extensions or contractions of the linkages to which the mass is attached. Elimination of the displacement variables from the resulting equations leads to values for μ and hence to the normal frequencies (Thompson and Healey 1936; Bartholomé and Teller 1932).

In the case of a non-linear molecule, in which, moreover, deformations of linkages are involved, the above procedure is less suitable, and it is customary and wiser to employ the alternative, though less visualizable, method. It is first necessary to write down an expression for the potential energy of the system when displaced from its rest position. Even after restricting the motion to being simple harmonic, several types of assumption can be made about this. We may, for example, suppose that there are forces between atoms joined by a conventional chemical bond, and also forces tending to prevent changes in the angles between any two valency bonds of the same atom. This may be called the "valency force system". For an isosceles molecule *ABA* with apex angle θ , the potential energy function can be written

$$V = \frac{1}{2}k_1\Delta_{AB}^2 + \frac{1}{2}k_1\Delta_{AB}^2 + \frac{1}{2}k_\theta\Delta\theta^2,$$

where k_1 is the force constant for elongation of the linkages $A-B$, and k_θ is the force constant for angular change in θ .

We may, on the other hand, assume that a force exists between any pair of atoms, regardless of whether a conventional valency bond exists. In this case, the "central force system", we write

$$V = \frac{1}{2}k_1\Delta_{AB}^2 + \frac{1}{2}k_1\Delta_{AB}^2 + \frac{1}{2}k_2\Delta_{AA}^2.$$

In more complex cases potential energy functions involving a greater number of constants can be set up, with a view to obtaining closer agreement with experimental data, but in such cases it is not infrequently found that the simpler expressions reproduce the data almost equally well (Howard

and Wilson 1934). It is obvious that factors specific to an individual molecule may make neither valency force field nor central force field strictly applicable, nevertheless, it seems to us preferable to use potential energy functions in which the force constants have some visualizable and precise physical meaning rather than those in which they have none. The facts of chemistry perhaps suggest a preference in this respect for valency force field rather than central force field. Penney and Sutherland (1936) have recently examined the relative merits of the two hypotheses, and have shown that experimental data favour the valency force field for the simple molecules which they have examined. They have also shown that the "cross-interaction terms" of the potential energy functions used are relatively insignificant and point out that in any case the experimental data do not justify the use of such more complex functions.

The normal frequencies of the isosceles triangular molecule ABA using valency force field have been calculated by Bjerrum (1914), by Yates (1930), and by Lechner (1932), and have since been restated by Penney and Sutherland (1936), and by Kohlrausch (1931). Since there appear to be several serious printing errors in the formulae given by Penney and Sutherland, whilst differences in nomenclature and in the definition of the fundamental constants make the expressions of the previous workers rather confusing, it is desirable to give the results of the calculations again here. The formulae, which will be needed below, will also serve as reference in connexion with further work shortly to be published on related chemical problems

For a molecule ABA , with atomic masses m_1 , m_2 , and m_1 , apex angle α , and link-length $A-B$ equal to d , and potential function described by equation (1), we find:

Valency Force Field

(a) Isosceles Triangular Model.

$$m_1^2 \nu_1^2 \nu_2^2 = \frac{2k_1 k_a}{d^2} \left\{ 1 + \frac{2m_1}{m_2} \right\},$$

$$m_1(\nu_1^2 + \nu_2^2) = k_1 \left\{ 1 + \frac{2m_1}{m_2} \cos^2 \alpha/2 \right\} + \frac{2k_a}{d^2} \left\{ 1 + \frac{2m_1}{m_2} \sin^2 \alpha/2 \right\},$$

$$\nu_3^2 = \frac{k_1}{m_1} \left\{ 1 + \frac{2m_1}{m_2} \sin^2 \alpha/2 \right\},$$

whence by elimination

$$k_a = \frac{m_1 \nu_1^2 \nu_2^2 d^2 (m_2 + 2m_1 \sin^2 \alpha/2)}{2(m_2 + 2m_1) \nu_3^2},$$

and for the angle α

$$x = x^2 \left\{ \frac{\nu_1^2 \nu_2^2}{\nu_3^2 (\nu_1^2 + \nu_2^2 + \nu_3^2) \left(1 + \frac{2m_1}{m_2} \right)} \right\} + \left\{ \frac{2\nu_2^2 \left(1 + \frac{m_1}{m_2} \right)}{(\nu_1^2 + \nu_2^2 + \nu_3^2)} \right\},$$

where

$$x = \left\{ 1 + \frac{2m_1}{m_2} \sin^2 \alpha/2 \right\}.$$

(b) *Linear Case.*

$$m_1^2 \nu_1^2 \nu_2^2 = \frac{2k_1 k_a}{d^2} \left\{ 1 + \frac{2m_1}{m_2} \right\},$$

$$m_1 (\nu_1^2 + \nu_2^2) = k_1 + \frac{2k_a}{d^2} \left\{ 1 + \frac{2m_1}{m_2} \right\},$$

$$\nu_3^2 = \frac{k_1}{m_1} \left\{ 1 + \frac{2m_1}{m_2} \right\},$$

whence by elimination

$$k_1 = m_1 \nu_1^2$$

and

$$k_a = \frac{m_1 m_2 \nu_2^2 d^2}{2(m_2 + 2m_1)}$$

or

$$k_a = \frac{m_1 \nu_1^2 \nu_2^2 d^2}{2\nu_3^2}.$$

ν_1 and ν_2 are the symmetrical vibrations, ν_3 the antisymmetrical one. The method of calculation (Whittaker 1927) is illustrated by the more complicated case discussed in Part IV below.

The corresponding formulæ for central force field, described by a potential energy function of equation (2), have been derived by Dennison (1926), by Radakowicz (1930, 1932), and others. They are:

Central Field Force

(a) *Isosceles Triangular Model.*

$$\nu_3^2 = \frac{k_1}{m_1} \left\{ 1 + \frac{2m_1}{m_2} \sin^2 \alpha/2 \right\},$$

$$m_1 (\nu_1^2 + \nu_2^2) = 2k_2 + k_1 \left\{ 1 + \frac{2m_1}{m_2} \cos^2 \alpha/2 \right\},$$

$$\nu_1^2 \nu_2^2 = \frac{2k_2}{m_1} \cdot \frac{k_1}{m_1} \cdot \left\{ 1 + \frac{2m_1}{m_2} \right\} \cos^2 \alpha/2,$$

whence

$$k_2 = \frac{m_1 m_2 \nu_1^2 \nu_2^2 \left(1 + \frac{2m_1}{m_2} \sin^2 \alpha/2 \right)}{2(m_2 + 2m_1) \nu_3^2 \cos^2 \alpha/2},$$

and for the angle

$$\begin{aligned} \sin^4 \alpha / 2 \left\{ \frac{4}{1 + \frac{2m_1}{m_2}} \cdot \frac{m_1^2}{m_2^2} \cdot \frac{\nu_1^2 \nu_2^2}{\nu_3^2} + \frac{2m_1}{m_2} (\nu_1^2 + \nu_2^2 + \nu_3^2) \right\} \\ + \sin^2 \alpha / 2 \left\{ \frac{4}{1 + \frac{2m_1}{m_2}} \cdot \frac{m_1}{m_2} \cdot \frac{\nu_1^2 \nu_2^2}{\nu_3^2} + \left(1 - \frac{2m_1}{m_2} \right) (\nu_1^2 + \nu_2^2 + \nu_3^2) - 2 \left(1 + \frac{m_1}{m_2} \right) \nu_3^2 \right\} \\ + \left\{ \frac{1}{1 + \frac{2m_1}{m_2}} \cdot \frac{\nu_1^2 \nu_2^2}{\nu_3^2} - (\nu_1^2 + \nu_2^2 + \nu_3^2) + 2 \left(1 + \frac{m_1}{m_2} \right) \nu_3^2 \right\} = 0. \end{aligned}$$

(b) *Linear Case.*

$$\nu_3^2 = \frac{k_1}{m_1} \cdot \left(1 + \frac{2m_1}{m_2} \right),$$

$$\nu_1^2 = \frac{2k_2}{m_1} + \frac{k_1}{m_1},$$

$$\nu_2^2 = 0.$$

It is seen that on this force field system the second (deformation) frequency of a linear molecule vanishes. This inherent weakness in the theory has been discussed by Radakowicz.

3—THE VIBRATION FREQUENCIES AND STRUCTURE OF ZINC DIMETHYL AND MERCURY DIMETHYL

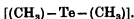
A—Zinc Dimethyl

The Raman frequencies of this molecule in the liquid state, given by Gopala Pai, and in agreement with the earlier values of Venkateswaran (1930), are:

$$144(2) \quad 248(1) \quad 488(2) \quad 505(8) \quad 620(4) \quad 1159(6) \quad 1346(0) \quad 2897(3).$$

It is very probable that the three largest frequencies involve vibrations in the methyl groups, and we are not at present concerned with them. The smaller frequencies involve vibrations of the central heavier atom. Using the assignment $\nu_1 = 620$, $\nu_2 = 144$, and $\nu_3 = 505$, and the formulae of the valency force field, Gopala Pai calculated for the apex angle α a value of 136° . As already explained, our measurements on the ultra-violet absorption spectra suggested the alternative assignment $\nu_1 = 505$, $\nu_2 = 144$, and $\nu_3 = 620$. Apart from the arguments which were advanced at that time, there

are several objections to Pai's assignment. First, we should expect the symmetrical frequency ν_1 to be the most intense in the Raman effect, whereas according to Pai the antisymmetrical frequency is the strongest. Secondly, all data on the dipole moments of this and similar compounds suggest a value of zero, i.e. linear structure with $\alpha = 180^\circ$, and measurements on electron diffraction confirm this (Braune and Knoke 1933). Thirdly, crystallographic evidence (Powell and Crowfoot 1932, 1934) has shown that the positive ion of thallium dimethyl iodide— $[\text{Te}(\text{CH}_3)_2]^+$ —which is iso-electronic with mercury dimethyl, has a linear arrangement



Denoting Gopala Pai's assignment as G.P. and ours as T.L., we can calculate the angle α in four ways:

- (1) Using central force field and G.P.
- (2) Using central force field and T.L.
- (3) Using valency force field and G.P.
- (4) Using valency force field and T.L.

(1) gives two roots, $\alpha = 136^\circ$, and a second imaginary value $\sin \alpha/2 \gg 1$; (2) gives two imaginary values; (3) gives an impossible value for α ; and (4) gives $\sin \alpha/2 = 1.02$, i.e. $\alpha = 180^\circ$.

Further, for linear triatomic molecules with valency force field,

$$\nu_1/\nu_3 = \sqrt{(m_2/m_1 + 2m_2)}.$$

With T.L. assignment, $\nu_1/\nu_3 = 0.814$ and $\sqrt{(m_2/m_1 + 2m_2)} = 0.82$, in good agreement

It seems certain, therefore, that valency force field and a linear structure with assignment $\nu_1 = 505$, $\nu_2 = 144$, and $\nu_3 = 620$ are the only means of satisfactorily explaining all the data, and it also brings the Raman intensities into line with what would be expected from symmetry considerations. That the antisymmetrical frequency should appear at all in the Raman effect may be surprising, but as Ingold and others (1936) have shown recently, selection rules in the Raman effect may break down with liquids where disturbing forces are liable to exist.

We can now calculate the force constants for zinc dimethyl using the valency force field formulae, and $m_1 = 15$, $m_2 = 65$, $\alpha = 180^\circ$ and $d = 2.08 \times 10^{-8}$ cm. (Sidgwick 1933). The values found are:

$$\begin{aligned} k_1 &= 2.31 \times 10^8 & \text{or} & \quad 2.24 \times 10^6 \text{ dynes/cm.}, \\ k_\alpha &= 0.27 \times 10^{-11} & \text{or} & \quad 0.26 \times 10^{-11} \text{ dyne-cm./radian.} \end{aligned}$$

The Raman frequencies can then be interpreted as follows:

Value	Designation	Type	Intensity
144	ν_3	Symmetrical	Medium
248	?	—	Weak
488	$\nu_3 - \nu_1$	—	Medium
505	ν_1	Symmetrical	Strong
620	ν_2	Antisymmetrical	Weak

B - Mercury Dimethyl

This molecule has the following Raman frequencies below 1000 cm^{-1} , the only ones which concern us

$$156(2) \quad 255(0) \quad 515(8) \quad 565(1) \quad 700(3).$$

The strongest frequencies are 156, 515 and 700, and Pai assigned these to ν_2 , ν_3 , and ν_1 respectively. Our previous measurements on the ultra-violet spectrum suggested the alternative assignment $\nu_1 = 515$, $\nu_2 = 156$, $\nu_3 = 700$. This appears at first sight to be much more preferable to that of Pai, in that, apart from other considerations, it gives 515 as the symmetrical and intense Raman frequency. A closer examination, however, reveals certain difficulties. We have again four alternatives in calculating the apex angle, namely,

- (1) Central force field with G.P. assignment
- (2) Central force field with T.P. assignment
- (3) Valency force field with G.P. assignment
- (4) Valency force field with T.L. assignment.

The values of α calculated in the different cases are

- (1) one value 130° , the second impossible, (2) two impossible values,
- (3) an impossible value; (4) an impossible value.

Thus the valency force field is impossible for either of the above assignments, and the central force field is clearly inadequate since it only gives a sensible value for α with a very unlikely assignment

It would be possible to argue that the difficulty arises from inherent shortcomings of the method of calculating the frequencies. We do not regard this as probable. It is unlikely that while zinc dimethyl can be so satisfactorily interpreted, the same argument should fail with mercury dimethyl.

Taking $\nu_1 = 515$, $\nu_2 = 156$, $\nu_3 = 565$, and assuming valency force field, we find $\sin \alpha/2 = 1.08$, i.e. $\alpha = 180^\circ$, as would be expected from other lines of work already referred to. Further, for the linear triatomic system $\nu_1/\nu_3 = 0.912$ and $\sqrt{(m_2/m_1 + 2m_2)} = 0.93$, in fairly good agreement. The only

difficulty about this assignment is the relative intensities of the different frequencies 565 and 700 given by Gopala Pai. Actually we should expect the unsymmetrical frequency to be weakly represented, and the real anomaly may be in the relative strength of this frequency with zinc dimethyl. It must be remembered however that both Raman lines involved are weak, and the intensity measurements of Gopala Pai are only approximate. The much heavier nature of the mercury central atom may also for some unforeseen reason lead to a higher intensity of the combination line 700.

We shall therefore assume for mercury dimethyl, $\nu_1 = 515$, $\nu_2 = 156$, and $\nu_3 = 565$, with $\alpha = 180^\circ$. Taking (Pauling and Huggins 1934) the value $d = 2.25 \times 10^{-8}$, we then find

$$k_1 = 2.33 \times 10^6 \quad \text{or} \quad 2.44 \times 10^5 \text{ dynes/cm.}$$

$$\text{and} \quad k_a = 0.46 \times 10^{-11} \quad \text{or} \quad 0.48 \times 10^{-11} \text{ dyne-cm /radian.}$$

The Raman lines of mercury dimethyl may then be interpreted as follows

Value	Designation	Type	Intensity
156	ν_2	Symmetrical	Weak
255	?	—	Very weak
515	ν_1	Symmetrical	Strong
565	ν_3	Antisymmetrical	Weak
700	$\nu_1 + \nu_2$	—	Medium

4—CALCULATION OF THE NORMAL VIBRATION FREQUENCIES OF ZINC DIETHYL AND MERCURY DIETHYL

Each of the above molecules can be regarded to a first approximation as a chain of five mass points $ABCBA$, in which on the basis of the above data the central BCB is 180° , and since the atoms B, B are essentially carbon atoms the angles $A\hat{B}C, C\hat{B}A$ are in the region of 109° . If free rotation is first regarded as impossible, the molecule may be expected to have two possible structures, either an inverted basin (fig. 2a) or a chair (fig. 2b).

Consider first the system as a whole. There will, in three-dimensional space, be $5 \times 3 = 15$ degrees of freedom. Of these, three will be reserved for translational motion and three for rotational motion, leaving nine for internal vibrations. In two dimensions there will be $[(5 \times 2) - (2 + 1)] = 7$ degrees of freedom for internal vibration. Thus seven of the nine proper vibrations will involve displacements of the masses in the plane of the molecule, and two will involve displacements perpendicular to the plane of the molecule.

Consider now the seven planar vibrations of the inverted basin form. These will in general split up into normal vibrations which are symmetric or antisymmetric to the symmetry element. The latter can be regarded as a plane through C perpendicular to the plane of the molecule or as a twofold rotation axis through C perpendicular to BCB and in the plane of the molecule. Let the displacements of the mass points A, B, C, B, A be taken as $x_1y_1x_2y_2x_3y_3x_4y_4x_5y_5$. We then have ten variables. For the symmetric vibrations, however, it can be seen that five simplifying conditions exist, namely $x_1 = -x_5, x_2 = -x_4, y_1 = y_5, y_2 = y_4, x_3 = 0$. In addition there will be a condition $\sum m_i y_i = 0$ for no translational motion. Thus we have ten variables

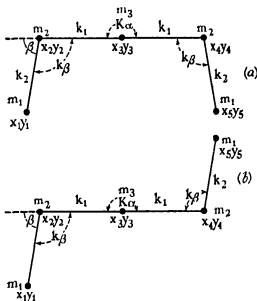


FIG. 2

and six simplifying conditions, and there will be four unknowns. This means that there will be four proper symmetric planar vibrations of the basin form. In the same way for the antisymmetric vibration there will be ten variables and seven simplifying conditions (one extra condition to eliminate rotation) giving three roots. Thus we have split up the group of nine proper vibrations of the basin into four planar symmetric, three planar antisymmetric, and two perpendicular to the plane of the molecule. Of the last two vibrations, one might be expected to be symmetric and the other antisymmetric, but we need not at present consider them further.

Similar considerations may be applied to the chair form (fig. 2b). Here there is a centre of symmetry C , and the system operation may be regarded

as a reflexion about this point, or as a rotation about a twofold rotation axis through C perpendicular to the plane of the molecule. The group of nine normal vibrations splits up into two vibrations perpendicular to the plane of the molecule, and seven planar vibrations; of these seven, three are symmetric and four antisymmetric.

The analysis of the group of nine normal frequencies in the above way simplifies the subsequent calculations. Four cases may now be considered separately: (a) the symmetrical basin, (b) the antisymmetrical basin, (c) the symmetrical chair, (d) the antisymmetrical chair. The calculation of (a) will now be given in detail and the results for the other cases summarized.

Case (a)—Symmetrical Basin Frequencies

Let the masses be m_1, m_2, m_3, m_2, m_1 , and the displacements of the particles $x_1, y_1, x_2, y_2, x_3, y_3, x_2, y_2, x_1, y_1$. Then for symmetric vibrations,

$$x_1 = -x_5, \quad (1)$$

$$x_2 = -x_4, \quad (2)$$

$$y_1 = y_5, \quad (3)$$

$$y_2 = y_4, \quad (4)$$

$$x_3 = 0. \quad (5)$$

These symmetry conditions automatically preclude translational motion along the x axis. For motion along the y axis, we can write, using equations (3) and (4),

$$2m_1y_1 + 2m_2y_2 + m_3y_3 = 0. \quad (6)$$

The kinetic energy

$$T = m_1\dot{x}_1^2 + m_1\dot{y}_1^2 + m_2\dot{x}_2^2 + m_2\dot{y}_2^2 + \frac{1}{2}m_3\dot{y}_3^2 \quad (7)$$

and the potential energy

$$V = 2\{\frac{1}{2}k_2\Delta_{12}^2 + \frac{1}{2}k_1\Delta_{23}^2 + \frac{1}{2}k_\beta\Delta\beta^2\} + \frac{1}{2}k_\alpha\Delta\alpha^2. \quad (8)$$

k_1 is the force constant of the $B-C$ link, k_2 that of $A-B$; β is the exterior angle of ABC , α the angle \hat{BCB} (180°).

We also have the geometric conditions

$$\Delta_{12} = \overline{x_1 - x_2} \cos \beta + \overline{y_1 - y_2} \sin \beta,$$

$$\Delta_{23} = \overline{x_2 - x_3} = x_2,$$

$$-d_2\Delta\beta = \overline{x_1 - x_2} \sin \beta - \overline{y_1 - y_2} \cos \beta + d_2/d_1 \cdot \overline{y_2 - y_3},$$

$$d_1\Delta\alpha = 2(y_2 - y_3).$$

Substituting these values into the potential energy function and applying the Lagrange equation,

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{x}_i} \right) = - \frac{\partial V}{\partial x_i},$$

and remembering that if the vibrations are to be simple harmonic $\ddot{x}_i = -\mu x_i$,

where $\nu = \frac{1}{2\pi} \sqrt{\mu}$, we have

$$\begin{aligned} x_1 \left\{ -k_2 \cos^2 \beta - \frac{k_\beta}{d_2^2} \sin^2 \beta + \mu m_1 \right\} \\ + y_1 \left\{ k_2 \cos^2 \beta + \frac{k_\beta}{d_2^2} \sin^2 \beta \right\} + y_1 \left\{ -k_2 \sin \beta \cos \beta + \frac{k_\beta}{d_2^2} \sin \beta \cos \beta \right\} \\ + y_2 \left\{ k_2 \sin \beta \cos \beta - \frac{k_\beta}{d_2^2} \sin \beta \cos \beta - \frac{k_\beta}{d_1 d_2} \sin \beta \right\} + y_3 \left\{ \frac{k_\beta}{d_1 d_2} \sin \beta \right\} = 0, \quad (9) \end{aligned}$$

$$\begin{aligned} x_1 \left\{ k_2 \cos^2 \beta + \frac{k_\beta}{d_2^2} \sin^2 \beta \right\} + x_2 \left\{ -k_2 \cos^2 \beta - \frac{k_\beta}{d_2^2} \sin^2 \beta - k_1 + \mu m_2 \right\} \\ + y_1 \left\{ k_2 \sin \beta \cos \beta - \frac{k_\beta}{d_2^2} \sin \beta \cos \beta \right\} \\ + y_2 \left\{ -k_2 \sin \beta \cos \beta + \frac{k_\beta}{d_2^2} \sin \beta \cos \beta + \frac{k_\beta}{d_1 d_2} \sin \beta \right\} + y_3 \left\{ -\frac{k_\beta}{d_1 d_2} \sin \beta \right\} = 0, \quad (10) \end{aligned}$$

$$\begin{aligned} x_1 \left\{ -k_2 \sin \beta \cos \beta + \frac{k_\beta}{d_2^2} \sin \beta \cos \beta \right\} + x_2 \left\{ k_2 \sin \beta \cos \beta - \frac{k_\beta}{d_2^2} \sin \beta \cos \beta \right\} \\ + y_1 \left\{ -k_2 \sin^2 \beta - \frac{k_\beta}{d_2^2} \cos^2 \beta + \mu m_1 \right\} \\ + y_2 \left\{ k_2 \sin^2 \beta + \frac{k_\beta}{d_2^2} \cos^2 \beta + \frac{k_\beta}{d_1 d_2} \cos \beta \right\} + y_3 \left\{ -\frac{k_\beta}{d_1 d_2} \cos \beta \right\} = 0, \quad (11) \end{aligned}$$

and

$$\begin{aligned} x_1 \left\{ k_2 \sin \beta \cos \beta - \frac{k_\beta}{d_2^2} \sin \beta \cos \beta - \frac{k_\beta}{d_1 d_2} \sin \beta \right\} \\ + x_2 \left\{ -k_2 \sin \beta \cos \beta + \frac{k_\beta}{d_2^2} \sin \beta \cos \beta + \frac{k_\beta}{d_1 d_2} \sin \beta \right\} \\ + y_1 \left\{ k_2 \sin^2 \beta + \frac{k_\beta}{d_2^2} \cos^2 \beta + \frac{k_\beta}{d_1 d_2} \cos \beta \right\} \\ + y_2 \left\{ -k_2 \sin^2 \beta - \frac{k_\beta}{d_2^2} \cos^2 \beta - \frac{2k_\beta}{d_1 d_2} \cos \beta - \frac{k_\beta}{d_1^2} - \frac{2k_a}{d_1^2} + \mu m_2 \right\} \\ + y_3 \left\{ \frac{k_\beta}{d_1 d_2} \cos \beta + \frac{k_\beta}{d_1^2} + \frac{2k_a}{d_1^2} \right\} = 0. \quad (12) \end{aligned}$$

Combining (9), (10), (11) and (12) with the translational condition (6) we arrive at the following determinantal equation:

$$\begin{vmatrix} A_1 & B_1 & C_1 & D_1 & E_1 \\ p_1\mu + A_2 & B_2 & C_2 & D_2 & E_2 \\ A_3 & p_2\mu + B_3 & C_3 & D_3 & E_3 \\ A_4 & B_4 & p_3\mu + C_4 & D_4 & E_4 \\ A_5 & B_5 & C_5 & p_4\mu + D_5 & E_5 \end{vmatrix} = 0,$$

in which

$$A_1 = B_1 = 0,$$

$$C_1 = 2m_1, \quad p_1 = p_3 = m_1,$$

$$D_1 = 2m_2, \quad p_2 = p_4 = m_2,$$

$$E_1 = m_3,$$

$$A_2 = -B_2 = -A_3 = (B_3 + k_1) = \left(-k_2 \cos^2 \beta - \frac{k_\beta}{d_2^2} \sin^2 \beta \right),$$

$$\begin{aligned} A_4 = -B_4 &= \left(-A_5 - \frac{k_\beta}{d_1 d_2} \sin \beta \right) = \left(B_5 - \frac{k_\beta}{d_1 d_2} \sin \beta \right) \\ &= \left(-k_2 \sin \beta \cos \beta + \frac{k_\beta}{d_2^2} \sin \beta \cos \beta \right), \end{aligned}$$

$$C_2 = -C_3 = A_4 = \left(-D_3 - \frac{k_\beta}{d_1 d_2} \sin \beta \right) = \left(D_3 - \frac{k_\beta}{d_1 d_2} \sin \beta \right),$$

$$C_4 = \left(D_4 - \frac{k_\beta}{d_1 d_2} \cos \beta \right) = \left(-C_5 + \frac{k_\beta}{d_1 d_2} \cos \beta \right) = \left(-k_2 \sin^2 \beta - \frac{k_\beta}{d_2^2} \cos^2 \beta \right),$$

$$D_5 = \left(C_4 - \frac{2k_\beta}{d_1 d_2} \cos \beta + \frac{k_\beta}{d_1^2} + \frac{2k_2}{d_1^2} \right),$$

$$E_2 = -E_3 = \frac{k_\beta}{d_1 d_2} \sin \beta,$$

$$E_4 = -\frac{k_\beta}{d_1 d_2} \cos \beta = -E_5 + \frac{k_\beta}{d_1^2} + \frac{2k_2}{d_1^2}.$$

Solving this, we find, writing $\frac{1}{M_2} = \left(\frac{1}{m_1} + \frac{1}{m_2} \right)$ and $\frac{1}{M_1} = \left(\frac{1}{m_3} + \frac{2}{m_2} \right)$,

$$\begin{aligned}
& \mu^4 \\
& - \mu^2 \left\{ k_1 \left(\frac{1}{m_2} \right) + k_2 \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_2^2} \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_1^2} \left(\frac{1}{M_1} \right) + \frac{k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \right) + \frac{2k_\alpha}{d_1^2} \left(\frac{1}{M_1} \right) \right\} \\
& + \mu^2 \left\{ k_1 k_2 \left(\frac{1}{m_2 M_2} - \frac{\cos^2 \beta}{m_2^2} \right) + \frac{2k_1 k_\alpha}{d_1^2} \left(\frac{1}{m_2 M_1} \right) + \frac{k_1 k_\beta}{d_1^2} \left(\frac{1}{m_2 M_1} \right) \right. \\
& \quad + \frac{k_1 k_\beta}{d_2^2} \left(\frac{1}{m_2 M_2} - \frac{\sin^2 \beta}{m_2^2} \right) + \frac{k_1 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2^2} \right) + \frac{k_2 k_\beta}{d_2^2} \left(\frac{1}{M_2} \right) + \frac{k_2 k_\beta}{d_1^2} \left(\frac{1}{M_1 M_2} - \frac{\sin^2 \beta}{m_2^2} \right) \\
& \quad \left. + \frac{k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_2} \right) + \frac{2k_2 k_\alpha}{d_1^2} \left(\frac{1}{M_1 M_2} - \frac{\sin^2 \beta}{m_2^2} \right) + \frac{2k_\alpha k_\beta}{d_1^2 d_2^2} \left(\frac{1}{M_1 M_2} - \frac{\cos^2 \beta}{m_2^2} \right) \right\} \\
& - \mu \left\{ \frac{k_1 k_2 k_\beta}{d_2^2} \left(\frac{1}{m_2 M_2} - \frac{1}{m_2^2 M_2} \right) + \frac{k_1 k_2 k_\beta}{d_1^2} \left(\frac{1}{m_2 M_1 M_2} - \frac{1}{m_2^2} \left[\frac{1}{m_2} + \frac{2 \cos^2 \beta}{m_3} \right] \right) \right. \\
& \quad + \frac{k_1 k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \left[\frac{1}{m_2 M_2} - \frac{1}{m_2^2} \right] \right) + \frac{2k_1 k_2 k_\alpha}{d_1^2} \left(\frac{1}{m_2 M_1 M_2} - \frac{1}{m_2^2} \left[\frac{1}{m_2} + \frac{2 \cos^2 \beta}{m_3} \right] \right) \\
& \quad + \frac{2k_1 k_\alpha k_\beta}{d_1^2 d_2^2} \left(\frac{1}{m_2 M_1 M_2} - \frac{1}{m_2^2} \left[\frac{1}{m_2} + \frac{2 \sin^2 \beta}{m_3} \right] \right) + \frac{2k_2 k_\alpha k_\beta}{d_1^2 d_2^2} \left(\frac{1}{M_1 M_2} - \frac{1}{m_2^2 M_2} \right) \Big\} \\
& + \left\{ \frac{2k_1 k_2 k_\alpha k_\beta}{d_1^2 d_2^2} \left[\left(\frac{1}{m_2 M_2} - \frac{1}{m_2^2} \right) \left(\frac{1}{M_1 M_2} - \frac{1}{m_2^2} \right) \right] \right\} = 0.
\end{aligned}$$

This equation gives the four values of μ which on insertion into the expression $\frac{1}{2\pi} \sqrt{\mu}$ give the four normal frequencies. If we now, knowing all the constants involved, insert the values of μ into the equations (6), (9), (10), (11), (12), and assume that $x_1 = 1$ unit, we may deduce the directions and relative magnitudes of the various atomic displacements.

Case (b)—Antisymmetrical Basin Frequencies

An entirely similar procedure is followed here, the only difference being that the symmetry conditions are $x_1 = x_3$, $x_2 = x_4$, $y_1 = -y_3$, $y_2 = -y_4$, $y_3 = 0$, and there is both a condition for no translation

$$m_3 x_3 + 2m_1 x_1 + 2m_2 x_2 = 0,$$

and for no rotation,

$$m_1 d_2 \sin \beta \cdot x_1 - m_1 (d_1 + d_2 \cos \beta) y_1 - m_2 d_1 y_2 = 0.$$

Using the same abbreviations as before, the solution is

$$\begin{aligned} & \mu^3 \\ & -\mu^3 \left(k_1 \left(\frac{1}{M_1} \right) + k_2 \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_2^3} \left(\frac{1}{M_3} \right) + \frac{k_\beta}{d_1^3} \left(\frac{1}{m_2} \right) + \frac{k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \right) \right) \\ & + \mu \left(k_1 k_2 \left(\frac{1}{M_1 M_2} - \frac{\cos^2 \beta}{m_2^3} \right) + \frac{k_1 k_\beta}{d_2^3} \left(\frac{1}{M_1 M_2} - \frac{\sin^2 \beta}{m_2^3} \right) + \frac{k_1 k_\beta}{d_1^3} \left(\frac{1}{M_1 m_2} \right) \right. \\ & \quad \left. + \frac{k_1 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_1} \right) + \frac{k_2 k_\beta}{d_2^3} \left(\frac{1}{M_2^3} \right) + \frac{k_2 k_\beta}{d_1^3} \left(\frac{1}{m_2 M_2} - \frac{\sin^2 \beta}{m_2^3} \right) + \frac{k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_2} \right) \right) \\ & - \left(\frac{k_1 k_2 k_\beta}{d_2^3} \left(\frac{1}{M_2^2 M_1} - \frac{1}{M_2 m_2^2} \right) + \frac{k_1 k_2 k_\beta}{d_1^3} \left(\frac{1}{m_2 M_1 M_2} - \frac{1}{m_2^3} - \frac{2 \sin^2 \beta}{m_2^3 m_2} \right) \right. \\ & \quad \left. + \frac{k_1 k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \left[\frac{1}{M_2 M_1} - \frac{1}{m_2^2} \right] \right) \right) = 0. \end{aligned}$$

Case (c)—Symmetrical Chair Frequencies

Here the symmetry conditions are $x_1, -x_3, x_2 = -x_4, y_1 = -y_5, y_2 = -y_4$, and $x_3 = y_3 = 0$. There is no independent condition for no translation but one for no rotation as in case (b). The solution is

$$\begin{aligned} & \mu^3 \\ & -\mu^3 \left(k_1 \left(\frac{1}{m_2} \right) + k_2 \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_2^3} \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_1^3} \left(\frac{1}{m_2} \right) + \frac{k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \right) \right) \\ & + \mu \left(k_1 k_2 \left(\frac{1}{m_2 M_2} - \frac{\cos^2 \beta}{m_2^3} \right) + \frac{k_1 k_\beta}{d_2^3} \left(\frac{1}{m_2 M_2} - \frac{\sin^2 \beta}{m_2^3} \right) + \frac{k_1 k_\beta}{d_1^3} \left(\frac{1}{m_2^3} \right) \right. \\ & \quad \left. + \frac{k_1 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2^3} \right) + \frac{k_2 k_\beta}{d_2^3} \left(\frac{1}{M_2^3} \right) + \frac{k_2 k_\beta}{d_1^3} \left(\frac{1}{m_2 M_2} - \frac{\sin^2 \beta}{m_2^3} \right) + \frac{k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_2} \right) \right) \\ & - \left(\frac{k_1 k_2 k_\beta}{d_2^3} \left(\frac{1}{m_2 M_2^2} - \frac{1}{m_2^2 M_2} \right) + \frac{k_1 k_2 k_\beta}{d_1^3} \left(\frac{1}{m_2^2 M_2} - \frac{1}{m_2^3} \right) \right. \\ & \quad \left. + \frac{k_1 k_2 k_\beta}{d_1 d_2} \left(\left[\frac{1}{m_2 M_2} - \frac{1}{m_2^2} \right] \frac{2 \cos \beta}{m_2} \right) \right) = 0. \end{aligned}$$

Case (d)—Antisymmetrical Chair Frequencies

In this case $x_1 = x_5, x_2 = x_4, y_1 = y_5, y_2 = y_4$ and x_3 and $y_3 = 0$. These relationships preclude rotation, but there are two conditions for no rotation, namely,

$$2m_1 x_1 + 2m_2 x_2 + m_3 x_3 = 0,$$

$$2m_1 y_1 + 2m_2 y_2 + m_3 y_3 = 0.$$

The solution is

$$\begin{aligned}
 & \mu^4 \\
 & -\mu^3 \left\{ k_1 \left(\frac{1}{M_1} \right) + k_2 \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_2^2} \left(\frac{1}{M_2} \right) + \frac{k_\beta}{d_1^2} \left(\frac{1}{M_1} \right) + \frac{k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \right) + \frac{2k_\alpha}{d_1^2} \left(\frac{1}{M_1} \right) \right\} \\
 & + \mu^2 \left\{ k_1 k_2 \left(\frac{1}{M_2 M_1} - \frac{\cos^2 \beta}{m_2^2} \right) + \frac{k_1 k_\beta}{d_2^2} \left(\frac{1}{M_2 M_1} - \frac{\sin^2 \beta}{m_2^2} \right) \right. \\
 & \quad + \frac{k_1 k_\beta}{d_1^2} \left(\frac{1}{M_1^2} \right) + \frac{k_1 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_1} \right) + \frac{2k_1 k_\alpha}{d_1^2} \left(\frac{1}{M_1^2} \right) + \frac{k_2 k_\beta}{d_2^2} \left(\frac{1}{M_2^2} \right) \\
 & \quad + \frac{k_2 k_\beta}{d_1^2} \left(\frac{1}{M_2 M_1} - \frac{\sin^2 \beta}{m_2^2} \right) + \frac{k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2 M_2} \right) \\
 & \quad \left. + \frac{2k_2 k_\alpha}{d_1^2} \left(\frac{1}{M_2 M_1} - \frac{\sin^2 \beta}{m_2^2} \right) + \frac{2k_\alpha k_\beta}{d_1^2 d_2^2} \left(\frac{1}{M_2 M_1} - \frac{\cos^2 \beta}{m_2^2} \right) \right\} \\
 & - \mu \left\{ \frac{k_1 k_2 k_\beta}{d_2^2} \left(\frac{1}{M_2^2 M_1} - \frac{1}{m_2^2 M_2} \right) + \frac{k_1 k_2 k_\beta}{d_1^2} \left(\frac{1}{M_2 M_1^2} - \frac{1}{m_2^2 M_1} \right) \right. \\
 & \quad + \frac{k_1 k_2 k_\beta}{d_1 d_2} \left(\frac{2 \cos \beta}{m_2} \left[\frac{1}{M_2 M_1} - \frac{1}{m_2^2} \right] \right) \\
 & \quad + \frac{2k_1 k_2 k_\alpha}{d_1^2} \left(\frac{1}{M_2 M_1^2} - \frac{1}{m_2^2 M_1} \right) + \frac{2k_1 k_2 k_\beta}{d_1^2 d_2^2} \left(\frac{1}{M_2 M_1^2} - \frac{1}{m_2^2 M_1} \right) \\
 & \quad \left. + \frac{2k_2 k_\alpha k_\beta}{d_1^2 d_2^2} \left(\frac{1}{M_2^2 M_1} - \frac{1}{m_2^2 M_2} \right) \right\} \\
 & + \frac{2k_1 k_2 k_\alpha k_\beta}{d_1^2 d_2^2} \left\{ \left(\frac{1}{M_2 M_1} - \frac{1}{m_2^2} \right)^2 \right\} = 0.
 \end{aligned}$$

5—NUMERICAL VALUES OF THE VIBRATION FREQUENCIES CALCULATED IN § 4

In order to determine the numerical magnitudes of the vibration frequencies calculated above, we require to know, in addition to the masses and the angle β , values of the force constants k_1 , k_2 , k_α , and k_β , in each of the cases of zinc diethyl and mercury diethyl.

$(180^\circ - \beta)$ is a modified tetrahedral angle and a consideration of the data for propane, butane and other relevant compounds suggests that we may take it to be about 115° , i.e. β equals 65° . Actually, variations of β by $\pm 5^\circ$ do not materially affect the values of the frequencies obtained.

k_2 is essentially a C—C stretching force constant. The frequency of ethane (Kohlrausch and Koppl 1934; Kohlrausch 1936) determined

essentially by extension or contraction of the linkage $(\text{H}_3\text{C})-(\text{CH}_3)$ is 993 cm^{-1} . Regarding the CH_3 groups as point masses of 15 units, $k_s = 4.34 \times 10^5 \text{ dynes/cm}$. In reality, however, we are concerned with a $(\text{H}_3\text{C})-(\text{CH}_3)$ linkage, and it may be preferable to determine this force constant from the data for propane, $(\text{H}_3\text{C})-(\text{CH}_2)-(\text{CH}_3)$, regarding the latter as an isosceles triangular molecule and using the valency force field formulae given above. In this case we can also determine k_β , the bending constant of the $-\text{CH}_2-$ angle, which is also needed in the present problem. Taking $\nu_1 = 867$, $\nu_2 = 377$, $\nu_3 = 1055$, and $d = 1.54 \times 10^{-8} \text{ cm}$, we find the apex angle to be 116° , $k_s = 3.86 \times 10^5$ and $k_\beta = 0.81 \times 10^{-11} \text{ dyne cm/radian}$. Thus the stretching constant of the C—C linkage has decreased by about 10% in passing from ethane to propane. Kohlrausch and Koppl (1934) have determined the value of the $(\text{H}_3\text{C})-(\text{CH}_2)$ constant in butane, and give $k_s = 3.85 \times 10^5$, in excellent agreement.

k_1 , the stretching constant of the $\text{Zn}-(\text{CH}_3)$ linkage, has already been calculated from the data above with zinc dimethyl. It may well be that the value 2.3×10^5 obtained must be decreased by 10–15% also as the linkage becomes $\text{Zn}-(\text{CH}_2)$, to a value of say 2.0×10^5 .

k_α , the bending constant of the zinc valency angle, has also been found above to be 0.265×10^{-11} . The value of k_β determined from the butane data is 1.30×10^{-11} , with methylene iodide $k_\beta = 1.15 \times 10^{-11}$, with methylene bromide $k_\beta = 1.4 \times 10^{-11}$, whereas with propane $k_\beta = 0.81 \times 10^{-11}$. It is thus apparent that k_β may vary quite considerably in the different compounds.

To summarize, therefore, we should expect the stretching constants to have the following approximate values:

$$k_1 = 2.0 \times 10^5, \quad k_s = 3.8 \times 10^5.$$

The bending constants may first of all be taken as

$$k_\alpha = 0.265 \times 10^{-11}, \quad k_\beta = 0.8 \times 10^{-11},$$

but we may expect to find that these last two values need emendation.

Using these values of the force constants, some of the higher frequencies calculated agree roughly in magnitude with those observed but there are discrepancies in the values of the smaller frequencies. This means that the values taken for k_α and k_β require alteration, and we are naturally driven to search for sensible adjustments in these (and if necessary the other) force constants which will bring the calculated data more into line. In doing this we must be guided by several considerations:

- (1) The molecules may be present in either only one, or both, of the two

theoretically possible forms (basin and chair). Accordingly the group of Raman frequencies observed may arise in one of several ways, (a) all being due to symmetrical frequencies of one or other form, present together, (b) all arising from one form only, some involving symmetric and others antisymmetric vibrations, (c) as a collection of both symmetrical and antisymmetrical frequencies from both forms present together.

(2) If the molecules are present in the "chair" form at all, we might expect lines due to symmetrical chair frequencies to be strongest in the Raman effect.

Of the various possibilities mentioned under (1), (a) might be most expected, qualified to some extent by (c).

Taking for zinc diethyl the values $k_1 = 1.6 \times 10^5$, $k_2 = 3.8 \times 10^5$, $k_3 = 0.8 \times 10^{-11}$, $k_\beta = 0.5 \times 10^{-11}$, $m_1 = 15$, $m_2 = 14$, $m_3 = 65$, $\beta = 65^\circ$, $d_1 = 2.08 \times 10^{-8}$ and $d_2 = 1.54 \times 10^{-8}$, we find the following values for the frequencies.

Symmetrical basin

$$\nu'_1 = 115$$

$$\nu'_3 = 245$$

$$\nu'_4 = 475$$

$$\nu'_6 = 962$$

Antisymmetrical basin

$$\nu'_2 = 220$$

$$\nu'_5 = 510$$

$$\nu'_7 = 960$$

Symmetrical chair

$$\nu_2 = 170$$

$$\nu_4 = 470$$

$$\nu_6 = 945$$

Antisymmetrical chair

$$\nu'_1 = 125$$

$$\nu'_2 = 270$$

$$\nu'_5 = 540$$

$$\nu'_7 = 995$$

The relative amplitudes of the different mass points in each of these vibrations are given in Table I, taking x_1 as one unit, and fig. 3 shows diagrammatically the form of the different vibrations.

It is seen at once that the calculated and observed values agree closely, more so than might have been expected. The force constants could, of course, easily be chosen so as to reproduce three or four of the frequencies accurately, but unless the values chosen and the method of calculation were essentially sound, the agreement would be confined to these three or four frequencies. Actually, it extends over the entire range of values, and, as will now be shown, in other ways than mere magnitude. Thus the adopted adjustments in the values of the force constants can be regarded as justified.

The columns of Table II show (1) the observed Raman frequencies of zinc diethyl and the intensities given by Gopala Pai, (2) suggested agreement with the symmetrical frequencies of both molecular forms, (3) suggested

TABLE I

Molecular form	Vibration type	Magnitude	x_1	x_2	x_3	y_1	y_2	y_3
Basin	Symmetric	115	+1.0	+0.08	0	-0.65	-0.16	+0.34
Basin	Symmetric	245	+1.0	+0.48	0	+0.66	+0.78	-0.64
Basin	Symmetric	475	+1.0	-7.9	0	-1.15	+3.36	-0.91
Basin	Symmetric	962	+1.0	-1.38	0	+2.3	-2.46	+0.02
Chair	Symmetric	170	+1.0	+0.37	0	+0.20	+0.44	0
Chair	Symmetric	470	+1.0	-14.0	0	-3.2	+5.2	0
Chair	Symmetric	945	+1.0	-1.0	0	+2.25	-2.45	0
Basin	Antisymmetric	220	+1.0	-0.2	-0.39	+0.09	+0.59	0
Basin	Antisymmetric	510	+1.0	+10.0	-4.8	+3.5	-4.2	0
Basin	Antisymmetric	960	+1.0	-1.4	+0.14	+2.24	-2.43	0
Chair	Antisymmetric	125	+1.0	-0.28	-0.34	-0.88	-0.27	+0.52
Chair	Antisymmetric	270	+1.0	-0.06	-0.44	+0.46	+0.80	-0.56
Chair	Antisymmetric	540	+1.0	-17.4	+7.0	-3.3	+7.2	-1.6
Chair	Antisymmetric	995	+1.0	-1.37	+0.13	-2.28	+4.91	-1.06

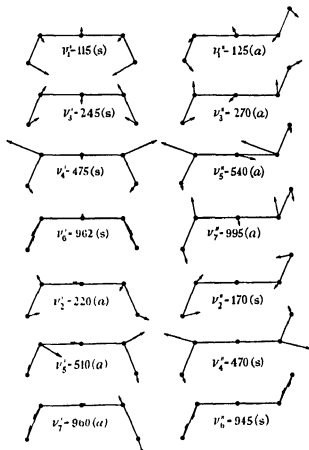


FIG. 3

agreements with the calculated "basin" frequencies, (4) suggested agreements with the calculated "chair" frequencies, and (5) probable final interpretation of the data.

TABLE II

(1)	(2)	(3)	(4)	(5)
111 (2)	115 (ν_1')	115 ($\nu_1' s$)	125 ($\nu_1'' a$)	115 ($\nu_1' s$)
176 (2)	170 (ν_2'')	220 ($\nu_2' a$)	170 ($\nu_2'' s$)	170 ($\nu_2'' s$)
255 (2)	245 (ν_3')	245 ($\nu_3' s$)	270 ($\nu_3'' a$)	245 ($\nu_3' s$)
476 (5)	475 (ν_4') 470 (ν_4'')	475 ($\nu_4' s$)	470 ($\nu_4'' s$)	475 ($\nu_4' s$) 470 ($\nu_4'' s$)
533 (1)	—	510 ($\nu_5 a$)	540 ($\nu_5' a$)	540 ($\nu_5' a$)
579 (2)	—	—	—	—
938 (2)	945 (ν_6'')	{ 960 } ($\nu_7' a$)	945 ($\nu_6'' s$)	945 ($\nu_6'' s$)
990 (2)	960 (ν_6')	{ 962 } ($\nu_8 s$)	995 ($\nu_7' a$)	995 ($\nu_7' a$)

Two general points relating to the magnitudes of the calculated frequencies should be noted. First, it is seen that the seven frequencies of the basin form are not very different in most cases from those of the chair. The relative form of the respective pairs of vibrations shown in fig. 3 makes this similarity understandable. Secondly, in the case of the basin form the close agreement between $\nu_1 = 960$ and $\nu_8 = 962$ may result in a resonance splitting to, say, 940–980.

The relative merits of the assignments implied in columns (2)–(5) can be summarized as follows:

In column (2) the agreement of the lowest frequencies is very satisfactory, all are symmetrical, as would be expected, and since both molecular forms give a symmetrical frequency of about 475, this might account for the marked intensity of this frequency in the Raman effect. The agreement amongst the higher frequencies is not so good. There is none to agree with the observed 533 and 579, and there is a wide discrepancy in the values 990 (observed) and 960 (calculated).

In column (3) the chief discrepancy is between the observed value of 176 and the calculated value of 220, although 510 (calculated) is rather far from 533 (observed). As before, there is no calculated value to explain the frequency observed at 579. This assignment explains satisfactorily the strong line at 476, and resonance splitting of the two frequencies 960–962 might well account for the observed pair 938–990. The main disadvantages to the scheme of column (3) is that it employs so many antisymmetrical vibrations.

The assignment used in column (4) is preferable to that of column (3), and might have been taken as an adequate explanation, were it not that this also involves the use of many antisymmetrical vibrations.

The assignment used in column (5) incorporates all the advantages of each of the other possibilities, and has no serious defects. All the frequencies agree in magnitude fairly well, and with the exception of the observed line at 579 all the Raman lines are accounted for. Moreover, all the calculated frequencies used involve symmetrical vibrations, except that at 533 (calculated 540) and this is, in fact, found to have a lower intensity. There is some doubt about the interpretation of the two observed frequencies 938–990. They may arise from a resonance doublet, but it is difficult to be more precise about this. The strength of the observed frequency 476 is fully in agreement with this scheme. It seems likely that the observed frequency 579 involves a non-planar vibration not calculated above. The calculation of the two non-planar frequencies associated with each molecular form, discussed in section III, is difficult, owing to inadequate knowledge of force constants, but a rough estimate indicates that with the basin form the two frequencies—each symmetrical—have values < 100 and *ca* 500 respectively. The first of these might well not have been detected by Gopala Pai

Using the data given above for mercury dimethyl and also the other relevant values previously taken, we might expect to use, for mercury diethyl, $k_1 = 2.1 \times 10^5$, $k_2 = 3.8 \times 10^5$, $k_\alpha = 0.47 \times 10^{-11}$ and $k_\beta = 0.8 \times 10^{-11}$. As before, it becomes necessary to use greatly amended values of the bending constants k_α and k_β , in order to bring the lowest frequencies into agreement. Using the values

$$k_1 = 1.7 \times 10^5, \quad k_2 = 3.9 \times 10^5, \quad k_\alpha = 1.7 \times 10^{-11}, \quad k_\beta = 0.6 \times 10^{-11}, \quad \beta = 63^\circ$$

and $d_1 = 2.25 \times 10^{-8}$, $d_2 = 1.54 \times 10^{-8}$,

the calculated frequencies are.

Symmetrical basin

$$\begin{aligned}\nu_1 &= 135 \\ \nu_2 &= 255 \\ \nu_4 &= 495 \\ \nu_6 &= 985\end{aligned}$$

Antisymmetrical basin

$$\begin{aligned}\nu_3 &= 215 \\ \nu_5 &= 505 \\ \nu_7 &= 970\end{aligned}$$

Symmetrical chair

$$\begin{aligned}\nu_2'' &= 200 \\ \nu_4'' &= 480 \\ \nu_6'' &= 970\end{aligned}$$

Antisymmetrical chair

$$\begin{aligned}\nu_1' &= 135 \\ \nu_3' &= 270 \\ \nu_5' &= 520 \\ \nu_7' &= 990\end{aligned}$$

The geometrical form of the different vibrations can be calculated in the same way as with zinc diethyl, and the results are in general not sensibly

different. Table III shows the frequencies set out in the same manner as in Table II.

TABLE III

(1)	(2)	(3)	(4)	(5)
140 (1)	135 (ν_1')	135 ($\nu_1' s$)	135 ($\nu_1'' a$)	135 ($\nu_1' s$)
212 (2)	200 (ν_2')	215 ($\nu_2 a$)	200 ($\nu_1' s$)	200 ($\nu_2' s$)
259 (3)	255 (ν_3')	255 ($\nu_3' s$)	270 ($\nu_3' a$)	255 ($\nu_3' s$)
329 (0)	—	—	—	—
480 (8)	$\begin{cases} 495 (\nu_4') \\ 480 (\nu_4') \end{cases}$	495 ($\nu_4' s$)	480 ($\nu_4' s$)	$\begin{cases} 495 (\nu_4' s) \\ 480 (\nu_4' s) \end{cases}$
502 (0)	—	505 ($\nu_5' a$)	520 ($\nu_5' a$)	520 ($\nu_5' a$)
633 (0)	—	—	—	—
958 (1)	970 (ν_6'')	970 ($\nu_7' a$)	970 ($\nu_6'' s$)	970 ($\nu_6'' s$)
1008 (3)	985 (ν_6'')	985 ($\nu_6' s$)	—	985 ($\nu_6' s$)
1055 (2)	—	—	990 ($\nu_7' a$)	990 ($\nu_7' a$)

In general the results are completely parallel with those seen for zinc diethyl. There are, however, certain small complications, and there is one large discrepancy between the calculated and observed frequency (562 obs. and 520 calc.) The observed frequency 633 is probably to be explained as with zinc diethyl as being due to a non-planar vibration. The same may be true of the feeble 329 vibration. The extra frequency in the region of 1000 may arise from more complex resonance splitting, for with the basin there are two close frequencies 985–970, and with the chair two more 970–990. Despite these minor differences, however, there is little doubt that in broad outline the observed data are fairly well reproduced.

6—GENERAL DISCUSSION

Perhaps the most important general conclusion to be drawn from the above is that the assumption of valency force field within molecules of the type discussed gives a fairly adequate description of the facts. In view of incomplete knowledge of the magnitude of the different force constants, and moreover, the variability of the constant for a given linkage in different compounds, it seems superfluous to set up a potential energy function involving a greater number of constants.

It can be seen that the bending constant k_s of the metal-carbon linkage is small, 0.26×10^{-11} for zinc-carbon in zinc dimethyl, and 0.47×10^{-11} for mercury-carbon in mercury dimethyl. The bending constant of the central atom angle in water is 0.61×10^{-11} and in hydrogen sulphide is 0.81×10^{-11} .

This means that the metal alkyl molecules are flexible, a result which has often been inferred from other lines of work.

It is also found above that in passing from the methyl derivatives to the ethyl compounds, whilst the values of the stretching constants k_1 and k_2 remain approximately the same, the values of k_a and k_β unexpectedly change considerably, and by the same percentage in the mercury compounds as in the zinc. Further, while k_a increases, k_β decreases. It might be natural to conclude from this that the electron distribution in the vicinity of the respective atoms involved has been changed. This suggests that an examination of the variation in different force constants in series of compounds may give a clue to the presence and mechanism of "electron drifts". In a later paper this will be more fully discussed. It will then be shown conclusively that alterations in the force constant of a given link in different compounds can be of profound significance in connexion with molecular structure, and the differences do not arise from inadequacies in the mathematical treatment.

Finally it is seen that the seven planar basin frequencies of the ethyl compounds are roughly equal in magnitude to the seven planar chair frequencies. This applies in particular to the frequencies 470 (symm. chair) and 475 (symm. basin) of zinc diethyl. Inspection of fig. 3 shows that if half of the chair form is rotated about the metal-carbon link so as to arrive at the basin type, the 470 vibration easily passes into the 475. We might infer from this that the normal frequency is the same "all the way round", and free rotation would not affect its magnitude.

We are grateful to the Government Grant Committee of the Royal Society, and to the Chemical Society, for grants, and one of us thanks the Department of Scientific and Industrial Research for a maintenance grant.

SUMMARY

The vibration frequencies of some metallic alkyl compounds have been considered in the light of different possible types of molecular force field. It is shown that with zinc dimethyl and mercury dimethyl, valency force field is much more satisfactory than central force field, and that experimental data are most consistent with an angle of 180° between the two valency bonds of the metallic atom in each case. Force constants for the linkages in these molecules have been calculated.

The general formulae for the normal vibration frequencies of molecules of the type of zinc diethyl and mercury diethyl have been calculated, and

using probable values for the appropriate force constants, their numerical magnitudes and geometric properties have been determined. The calculated results agree very well with experimental data.

The significance of some of the results found is discussed in connexion with chemical problems. The results suggest that a consideration in different series of compounds of the variation of force constants of links, for stretching and bending respectively, may give indications of localization of electrons or of electron drifts. This will be more fully examined in a later paper.

REFERENCES

- Bartholomé and Teller 1932 *Z. phys. Chem. B*, **19**, 366.
Bjerrum 1914 *Verh. phys.-med. Ges. Würzburg*, **16**, 737.
Braune and Knoke 1933 *Z. phys. Chem. B*, **23**, 163.
Dennison 1926 *Phil. Mag.*, **1**, 195.
Gopala Pai 1935 *Proc. Roy. Soc. A*, **149**, 33.
Herzberg and Teller 1933 *Z. phys. Chem. B*, **21**, 410.
Howard 1935 *J. Chem. Phys.*, **3**, 207.
Howard and Wilson 1934 *J. Chem. Phys.*, **2**, 630.
Ingold and others 1936 *J. Chem. Soc.* p. 986.
Kohlrausch 1931 "Der Sinek-Raman Effekt." pp 170-80. Leipzig.
— 1936 *S.B. Akad. Wiss. Wien*, **145**, 569.
Kohlrausch and Koppl 1934 *Z. phys. Chem. B*, **26**, 209.
Lochner 1932 *S.B. Akad. Wiss. Wien*, **141**, 291, 633.
Pauling and Huggins 1934 *Z. Kristallogr.*, **87**, 205.
Penney and Sutherland 1936 *Proc. Roy. Soc. A*, **156**, 654, 678.
Powell and Crowfoot 1932 *Nature, Lond.*, **130**, 131.
— — 1934 *Z. Kristallogr.*, **87**, 370.
Radakowicz 1930 *S.B. Akad. Wiss. Wien*, **139**, 107.
— 1932 *S.B. Akad. Wiss. Wien*, **141**, 41.
Sidgwick 1933 "The Covalent Link," p. 83. Cornell.
Sutherland 1936 *Nature, Lond.*, **138**, 641.
Sutherland and Dennison 1935 *Proc. Roy. Soc. A*, **148**, 250.
Thompson and Healey 1936 *Proc. Roy. Soc. A*, **157**, 331.
Thompson and Linnett 1934 *J. Chem. Soc.* p. 700.
— — 1935 *Proc. Roy. Soc. A*, **150**, 603.
— — 1936 *Proc. Roy. Soc. A*, **156**, 108.
— — 1937 *Trans. Faraday Soc.* (in the Press).
Venkateswaran 1930 *Indian J. Phys.*, **5**, 145.
Whittaker 1927 "Analytical Dynamics," chap. vii. Cambridge.
Yates 1930 *Phys. Rev.*, **36**, 555.
— 1931 *Phys. Rev.*, **37**, 716.
-

The Ionization Potentials of the Free Radicals Methyl and Ethyl

BY R. G. J. FRASER, *Imperial Chemical Industries, Limited*
AND T. N. JEWITT, *Clare College, Cambridge*

(Communicated by R. G. W. Norrish, F.R.S.—Received 16 January 1937)

INTRODUCTION

Molecular beams, that is beams of neutral molecules moving with thermal velocities *in vacuo*, possess two salient characteristics: narrow beams are, to a good approximation, unidirectional; and they are essentially collision free. In 1933 Fraser (1934) suggested that the collision-free character of molecular beams could be utilized in the study of free radicals, such as free methyl and free ethyl, which are the primary products of so many chemical reactions

This suggestion promised to be an immediately practicable one, inasmuch as Estermann and Stern (1933) had already found in the space-charge detector not only an extremely sensitive means of measuring the intensity of a molecular beam, but also one which looked to be applicable to the detection of beams of any molecular species. The detector is a diode, the cathode a tungsten wire threaded through the closed ends of the cylindrical anode; the ratio of the diameters of wire and cylinder being so chosen, for a given range of anode potentials, that the electron emission from the filament is governed by its negative space charge. The beam enters the diode through a small aperture, and establishes therein an equilibrium pressure, when as many molecules escape, by the beam aperture and filament clearance holes, as the beam brings in. If the potential difference between the electrodes is greater than the ionization potential of the entering molecules, positive ions will be formed, which partially neutralize the negative space charge and hence increase the electron emission from the filament. It is this increase of electron emission which is a measure of the intensity of the entering beam.*

The detector will not respond to the entry of a molecular beam, whatever its intensity, unless the difference of potential between the electrodes exceeds the ionization potential of the molecules constituting the beam;

* For details of the electrical circuits used, see Estermann and Stern (1933), or Jewitt (1934).

and Fraser (1934) suggested that in this fact might lie the basis of a method for the qualitative analysis of the primary products of chemical reactions. To this end it was essential to discover whether the molecular beam technique was generally capable of measuring ionization potentials with reasonable accuracy. Jewitt (1934) therefore made determinations of the ionization potentials of formaldehyde and of the methyl halides, relative to that of mercury as standard substance. Subsequently they proved to be in satisfactory agreement (within 0.5 V) with the measurements of Price (1936), who used a spectroscopic precision method.

We could therefore proceed with some confidence to the measurement of the ionization potentials of free methyl and free ethyl by the molecular beam method; data which are of intrinsic value, and at the same time the necessary preliminary to the identification of the free radicals in a qualitative analysis. It is these experiments which we shall describe in the present paper.*

OUTLINE OF METHOD

The essential parts of the apparatus used, all of which are within an evacuated envelope, are shown schematically in fig 1. The vapour of either

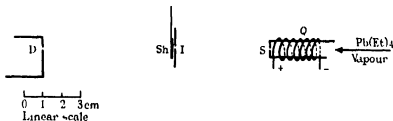


FIG. 1.—The Beam System.

tetra-methyl or tetra-ethyl lead enters a heated quartz tube Q , where the lead is deposited, and the corresponding free radicals are liberated (Paneth and Hofeditz 1929; Paneth and Lautsch 1931). The free radicals emerge from the source aperture S , and a beam of them is defined by the image aperture I . The electro-magnetically operated shutter Sh controls the entry of the beam into the detector D .

In preliminary experiments with lead tetra-ethyl, it was found that a beam of undecomposed lead tetra-ethyl (quartz tube unheated), entering a

* A preliminary account of some of our results has already been given in the *Physical Review*, December 1936.

space-charge detector at *D*, produced the normal increase in the electron emission for electron energies greater than about 12V. With a beam of free ethyl radicals (quartz tube heated), on the other hand, a progressive decline in the amount of the increase, and eventually a decrease in electron emission occurred. The most likely explanation appeared to be that the reactive free radical hydrocarbons give rise to carbide formation at the hot filament, with a consequent decrease in the value of the electron emission *in vacuo* sufficient to offset, or even outweigh, neutralization of the space charge caused by the positive ions. If this explanation is correct, the decrease in emission caused by carbide formation must be of the same order as the increase caused by the neutralization of the space charge, namely, some 10 μ A only in 3mA. It was therefore decided to replace the space-charge detector by a suitable ionization gauge, and to obtain an ionization curve by plotting the positive ion current produced by the beam against electron energy; when a decrease in electron emission of the amount to be expected from the assumed carbide formation would have a relatively small effect on the positive ion current in actual practice, decrease by a factor 10^{-4} as against increase by a factor of 10^{-2} due to ionization of the free radicals. It may be said at once that the decision to take this step justified itself completely.

THE APPARATUS

A schematized drawing of the apparatus in its final form is seen in fig. 2. It is divided into two main chambers, the collimator chamber (I) and the detector chamber (II), by the diaphragm carrying the 1 mm. diameter image aperture (6). The collimator chamber is pumped at (4) by a Metropolitan-Vickers Apiezon oil pump 03, having a speed of 20 l. μ ; this high speed is necessary, since the free radicals scattered at the walls of the collimator chamber combine to form the corresponding hydrocarbons, which are not readily condensable. The detector chamber is pumped at (7) by a mercury diffusion pump, Leybold Type B, having an effective speed of some 5 l. μ ; a mercury pump was chosen for the detector chamber as having a better end-vacuum than the oil diffusion pumps. In addition to the pumps, large liquid air-cooled surfaces at (5) and (8) were introduced to clean up undecomposed lead tetra-alkyl vapour. The pressures in the collimator chamber (3×10^{-5} mm.) and detector chamber (10^{-6} mm.) were measured at ionization gauges (not shown).

The vapours of both lead tetra-methyl and lead tetra-ethyl are extremely poisonous. Liquid air traps were therefore introduced into the fore-vacuum

line of both pumps; after a run the liquid air was removed from (5) and (8), and pumping continued long enough to ensure that as far as possible all traces of undecomposed tetra-alkyl lead had collected in the fore-vacuum traps. The traps were then removed, and the tetra-alkyl lead destroyed, by heating to 300° C. in a hot-air oven with an efficient draught.

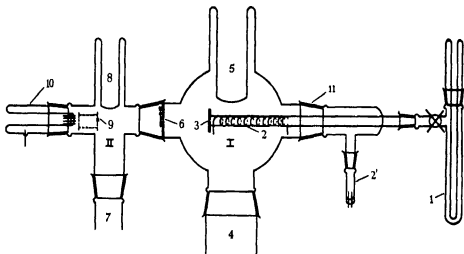


FIG. 2.—General View of Apparatus.

Source of Free Radicals. The tetra-alkyl leads, which are volatile liquids, were stored in the vessel (1), fitted with an internal tube carrying a thermocouple. The lead tetra-ethyl, obtained commercially, was previously purified by distillation under reduced pressure in a stream of nitrogen, with subsequent redistillation of the appropriate fraction. The lead tetra-methyl used required no previous treatment.*

The vessel (1) was surrounded by a freezing mixture, at a temperature (−10 to 0° C.) which ensured molecular streaming, either of the free radicals or the undecomposed vapour of the tetra-alkyl lead, from the source orifice (3), a circular hole 0.5 mm. diameter at the centre of a mild steel plate, held by a backing ring and three tension screws against the flanged end of the quartz tube (2). Knowing the speed of the 03 pump at (4), the pressure in the collimator chamber, and the area of the source orifice, the effective pressure behind the source orifice is easily estimated (Fraser 1931, p. 21).

* We are greatly indebted to Dr Aston for the gift of samples of pure lead tetra-methyl in sealed glass containers.

The value of the source pressure for molecular streaming at the orifice (3) is of the order 10^{-3} mm. An accurate knowledge of its absolute value is unnecessary in the present experiments, but it is important to keep track of fluctuations occurring during the determination of an ionization curve. For a given electron energy, the positive ion current depends not only on the ionization probability of the molecules entering the detector, but also on the intensity of the beam; and this is determined, for a given beam system, by the source pressure. A simple form of Pirani gauge (not shown) was therefore introduced into the tube connecting storage vessel and apparatus proper; and it was found that small variations in the source pressure were proportional with sufficient accuracy to the corresponding changes in the gauge wire current, when a constant potential difference was maintained between the ends of the wire. For any one substance, therefore, the positive-ion currents in the detector could all be referred to a constant pressure in the source. The gauge also gave a rough check on the equality of the source pressure with different substances.

The experience of Paneth and his co-workers was followed closely in ensuring that complete thermal decomposition of the tetra-alkyl lead to metallic lead and free alkyl radicals took place in the heated quartz tube (2).^{*} Thus the nichrome ribbon heating spiral, the lead-in for which is seen at (2'), was graduated from close winding where the vapour enters the quartz tube to more open spacing towards the source orifice, in order to suppress "self-eating" of the lead mirror. The appearance of the lead deposit is itself a rather safe criterion of complete decomposition (Paneth and Hofeditz 1929); as a control, however, a quantitative estimate of the lead deposited in the quartz tube was made in a series of test runs (cf. Paneth, Hofeditz, and Wunsch 1935, p. 373), and compared with the amount of tetra-alkyl lead lost from the storage vessel (1).

The Ionization Gauge Detector—The detector (9), essentially a standard triode gauge, is carried by a four wire pinch on the ground joint (10). It is shown in detail in fig. 3. The anode *A* is, except for the necessary lead and beam openings, a cylinder with closed ends, spot-welded from sheet nickel. It is held between the quartz tubes *QQ'*, let into V grooves in the flanges formed by the closed ends; the tubes are supported in their turn by the two outermost leads, which are threaded through them. These leads also serve as negative lead and one support for the filament *F*; the third

^{*} Paneth, as a matter of convenience, used hydrogen or helium as "carrier gas" in the majority of his experiments. Later Paneth, Hofeditz and Wunsch (1935, p. 378) showed the practicability of working without a carrier gas, as is of course necessary in our experiments.

lead from the pinch serves as positive lead and the second support for the filament, through the spring S_1 , which is correctly dimensioned in relation to the filament (Blodgett and Langmuir 1934). The fourth wire of the pinch serves as anode lead. The copper band G is connected to the negative filament lead and to earth; it forms a guard ring between the seals carrying the anode and filament and that through which the grid lead Gl enters the apparatus. The long lead carrying the grid Gr is supported near its centre by the quartz sleeve and spring S_2 . This construction has the following advantages. (1) rigidity, (2) accessibility of the filament when replacement is necessary; (3) good insulation (15 cm. of glass) between guard ring and grid, (4) a relatively high κ factor (0.2) (Fraser 1931, p. 34)

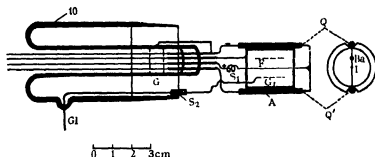


FIG. 3—Ionization Gauge Detector.

Alignment of the Beam System--Alignment was carried out optically. The storage vessel (1) was removed, the shutter opened, and a beam of light shone through source and image apertures on to the closed end of the anode. The anode is mounted excentrically to the axis of the joint (10), as is also the source tube on its joint (11); and by turning these joints the spot of light could be brought on to the 1.5 mm diameter beam aperture Ba (fig. 3).

METHOD OF MEASUREMENT

The positive ion current to the grid of the gauge was measured by means of an electrometer valve D.C. amplifier, in conjunction with a Leeds and Northrup moving coil galvanometer, used at a current sensitivity of 10^{-10} amp./mm. at 1 m. The valve was an Osram Type T, mutual transconductance 0.08 mA/V at anode potential +6V, thus with a grid leak of 10 ohms, the current sensitivity of the arrangement is 1.3×10^{-15} amp./mm. An advantage of the molecular beam method is the rapidly controllable admission to or exclusion from the detector of the substance being ex-

aminated; this makes possible a series of timed readings of the galvanometer deflexions, which serves at once to eliminate a slow change in the zero of the circuit and to average over random sudden fluctuations (Knauer and Stern 1929). Thus in the present experiments it was not essential to have a neutralized circuit, such as those described by Du Bridge and Brown (1933) or by Harnwell and Van Voorhis (1934), in order to obtain adequately reproducible results at high sensitivity.

Once vapour is admitted to the source, molecules of the beam substance are in uncoordinated motion in the observation chamber, whether the shutter is open or closed, since the image aperture acts also as a source aperture for the collimator chamber. These scattered molecules gain access to the detector, and produce a background to its responses to the entry of the beam. It is therefore standard practice to have a compensation gauge in the observation chamber, near the detector gauge but out of the path of the beam (Fraser 1931, p. 35). In the present case this was departed from, in order to avoid serious complication of the amplifier circuit; and instead the background positive ion current was balanced out with the main plate current of the electrometer valve, by means of the usual output potentiometer. The balance was made before each set of readings of the galvanometer deflexions for a given anode potential of the ionization detector, and checked thereafter. With this procedure, the average of a set of galvanometer deflexions was reproducible to within $\pm 10\%$.

Fig. 4 illustrates the method of determining the ionization potentials of the free radicals. The ionization curve of a standard substance, in the present experiments methyl iodide, is observed, and thereafter a similar curve for the free radical, here C₂H₅. The electron energies, as read on an accurate voltmeter, differ from the true electron energies by a constant amount, determined mainly by contact potentials in the apparatus; comparison of the ionization curve with that of a substance whose ionization potential is known serves to determine this constant, and hence a correct scale of electron energies is obtained. Thus in the present example the observed ionization potential of methyl iodide is 11.2V, its true ionization potential is 9.489V (Price 1936), the constant of correction is therefore 1.7V, and finally the true ionization potential of free ethyl is determined as $(12.3 - 1.7) = 10.6\text{V}$.

RESULTS

The ionization curves for free ethyl (source nozzle heated) and the parent lead tetra-ethyl (source nozzle unheated) are shown in figs. 5 and 6,

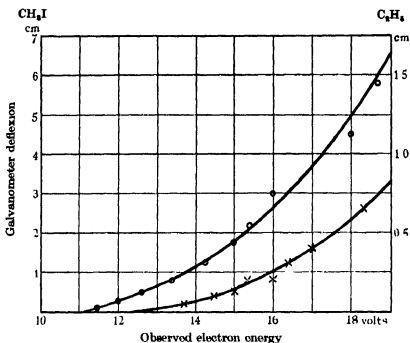


FIG. 4—Illustrating method of correction of the observed ionization potentials.

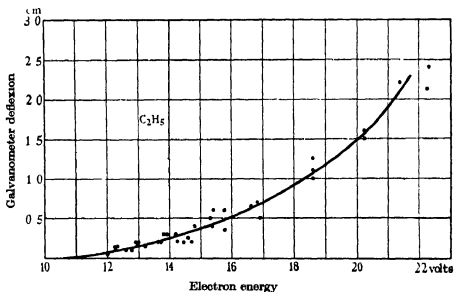


FIG. 5—Free ethyl: ionization potential, 10.6 V.

the scale of electron energies being the true scale arrived at in the way described above. In both figures, all the points taken in four separate runs are given equal weight.

Fig. 7 shows a representative series of measurements for free methyl and the parent lead tetra-methyl. The same degree of reproducibility was never

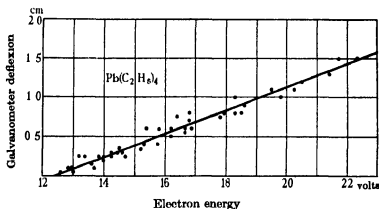


FIG. 6—Lead tetra-ethyl: ionization potential, 12.3 V.

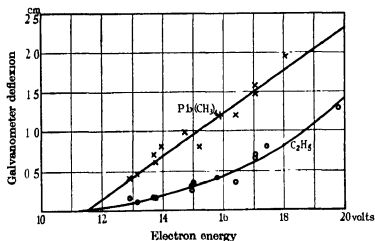


FIG. 7—Free methyl: ionization potential, 11.2 V (\circ); lead tetra-methyl: ionization potential (11.5 V) (\times).

attained with $\text{Pb}(\text{CH}_3)_4$ as with CH_3 , C_2H_5 , and $\text{Pb}(\text{C}_2\text{H}_5)_4$. Owing to the departure of both authors from Cambridge, it has not been possible to trace this trouble to its source, and the value (11.5 V) for the ionization potential of $\text{Pb}(\text{CH}_3)_4$ we therefore regard as provisional only.

The curves for free methyl and free ethyl are plotted together in fig. 8. They follow approximately parallel courses, as the analogous measurements with the methyl halides (Jewitt 1934) would lead one to expect.

Comparison of figs. 5, 6 and 7 shows that the ionization curves of the free radicals on the one hand and the tetra-alkyl leads on the other differ systematically in shape within the range of electron energies covered by the measurements. The linear upward break observed with the tetra-alkyl leads is exceptional for molecular ions.

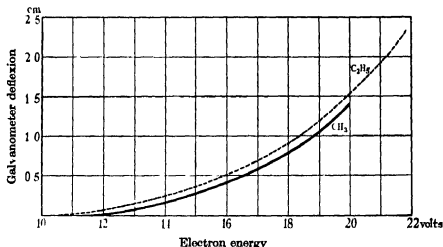


FIG. 8—Comparison of the ionization curves for free methyl (full line) and free ethyl (broken line).

The ionization potentials, obtained by extrapolation of the ionization curves to meet the axis of abscissae, are set out in Table I, the estimated error being ± 0.8 V.

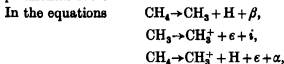
TABLE I

Substance	Ionization potential (V)
$\text{Pb}(\text{CH}_3)_4$	(11.5)
$\text{Pb}(\text{C}_2\text{H}_5)_4$	12.3 ± 0.8
Free methyl	11.2
Free ethyl	10.6 ± 0.8

DISCUSSION

The present measurements are, so far as we know, the first determinations by any method of the ionization potentials of free radicals. In the

case of free methyl, however, it is possible, with certain assumptions, to form an indirect estimate of its ionization potential on the basis of existing experimental data.



β denotes the strength of the C—H bond in methane, i the ionization potential of CH₃, α the appearance potential of CH₃⁺ when formed by the electron bombardment of methane; while ϵ denotes a freed electron.

Thus $i = \alpha - \beta$, if one assumes that no internal or kinetic energy is to be assigned to the products of the reaction $\text{CH}_4 \rightarrow \text{CH}_3^+ + \text{H}$, otherwise $i < \alpha - \beta$. β may be taken as 4.5 V (Norrish 1934), while α has been determined independently by Hogness and Kvalnes (1928) Hipple and Bleakney (1935) and Smith (1937) as 15.5, 14.7 and 14.4 V respectively. The corresponding values of the ionization potential i are summarized, in comparison with our directly determined value, in Table II.

TABLE II

Observers	Ionization potential (V)
Fraser and Jewitt	11.2 ± 0.8
Hogness and Kvalnes (1928)	≈ 11.0
Hipple and Bleakney (1935)	≈ 10.2
Smith (1937)	≈ 9.9 ± 0.4

Comparing first the directly determined value with the appearance potential data of Hogness and Kvalnes, some 0.6 V energy could be assigned to the products of the reaction $\text{CH}_4 \rightarrow \text{CH}_3^+ + \text{H}$; this is quite closely the value (0.68 V ± 10%) estimated by Penney (1935) for the energy of rearrangement of the unstable tetrahedral form of free methyl to the stable plane form. On the other hand, the directly determined value can only be brought into relation with the most recent appearance potential measurements, those of Smith, by assuming that the products of the above reaction are formed without energy.

Mulliken (1933) has calculated the ionization potential of free methyl as roughly 8.5 V. This value stands in very reasonable relationship to Smith's data, but is incompatible with our directly determined value. It would be of interest to have a more exact theoretical estimate.

The work reported was financed by Imperial Chemical Industries Limited, and was carried out in the Laboratory of Physical Chemistry,

Cambridge. We wish to record our appreciation of the late Professor Lowry's interest in the experiments, and of his unfailing helpfulness.

SUMMARY

The ionization potentials of free methyl and free ethyl are measured by a molecular beam method. The beam of free radicals, formed in the source by the thermal decomposition of the corresponding tetra-alkyl lead, is received at an ionization gauge, and curves relating positive ion current to electron energy are obtained. Extrapolation of these curves to meet the axis of abscissae determines the ionization potentials, the scale of electron energies being corrected against that of methyl iodide.

REFERENCES

- Blodgett and Langmuir 1934 *Rev. Sci. Instrum.* **5**, 321.
Du Bridge and Brown 1933 *Rev. Sci. Instrum.* **4**, 532.
Estermann and Stern 1933 *Z. Phys.* **85**, 135.
Fraser 1931 "Molecular Rays." Cambridge.
— 1934 *Trans. Faraday Soc.* **30**, 182.
Harnwell and Van Voorhis 1934 *Rev. Sci. Instrum.* **5**, 244.
Hipple and Bleakney 1935 *Phys. Rev.* **47**, 802.
Hogness and Kvalnes 1928 *Phys. Rev.* **32**, 942.
Jewitt 1934 *Phys. Rev.* **46**, 616.
Knauer and Stern 1929 *Z. Phys.* **53**, 766.
Mulliken 1933 *J. Chem. Phys.* **1**, 492.
Norrish 1934 *Trans. Faraday Soc.* **30**, 105.
Paneth and Hofeditz 1929 *Ber. dtsch. chem. Ges.* **62**, 1335.
Paneth and Lautsch 1931 *Ber. dtsch. chem. Ges.* **64**, 2702.
Paneth, Hofeditz and Wunsch 1935 *J. Chem. Soc.* p. 372.
Penney 1935 *Trans. Faraday Soc.* **31**, 734.
Price 1936 *J. Chem. Phys.* **4**, 539.
Smith 1937 *Phys. Rev.* **51**, 263.
-

Physical Properties of Surfaces

IV—Polishing, Surface Flow and the Formation of the Beilby Layer

BY F. P. BOWDEN AND T. P. HUGHES

Laboratory of Physical Chemistry, Cambridge

(Communicated by R. G. W. Norrish, F.R.S.—Received 11 February 1937)

[Plates 5–11]

The usual method of polishing surfaces is to rub them together with a fine powder between them. By this process a rough surface, having visible surface irregularities, is changed into one where the irregularities are invisible. If the surface gives specular reflexion the height of these irregularities will be less than half a wave-length of visible light. Beilby has shown that the top layer of the polished solid is "vitreous" in character, it has lost its obvious crystalline properties and has apparently flowed over the surface, bridging and filling up the irregularities in it. The mechanism of the process has been a subject of discussion for many years. Newton (1730, p. 265), Herschel (1830, p. 447) and Rayleigh (1901a) considered that polishing was essentially due to abrasion. Beilby (1921, p. 114) considered that the polisher tore off the surface atoms, and the layer below this "retains its mobility for an instant, and before solidification is smoothed over by the action of surface tension". Adam (1927) has suggested that molecules or small particles are abraded from the surface and subsequently adhere to a different area, so building up an amorphous layer. Hamburger (1932) considered that the abraded units were not molecules but small crystals perhaps 30 Å in diameter. The view that the temperature is sufficiently high to cause surface melting has been held for some time by Macaulay (1926, 1927, 1931), who was able to detect the products of thermal decomposition of the powder used for polishing glass plates. This evidence has been questioned because many reactions may occur at freshly exposed surfaces due to effects other than temperature. J. W. French (1927) carried out experiments with a thermometer embedded in the polisher, and concluded that the rise in surface temperature was negligibly small.

Experiments described in earlier papers (Bowden and Ridler 1935, 1936) provide direct evidence that the surface temperature at the points of contact may, under many conditions of sliding, be sufficiently high to cause a real melting of the metal. The surface temperature was determined by

using the rubbing contact of two different metals as a thermo-couple and measuring the e.m.f. developed on sliding. The mass of the metal remained quite cool, the high temperature was localized at the surface of the points of contact. It is just at these points, however, that wear, surface flow and polish occur. The method can, of course, only be applied to metallic conductors, but experiment showed that there was a simple relationship between surface temperature and thermal conductivity. A metal of low thermal conductivity, such as bismuth, showed a much higher surface temperature than copper under the same conditions of sliding. In the case of non-metals such as glass, silk, alumina, etc., which possess a very low thermal conductivity, we should expect the local surface temperature to be very much higher still. If these high local temperatures are reached so easily when one surface is rubbed on another, we should expect it to play an important part in the mechanism of wear, surface flow and polish of solids. The melted or softened solid would be wiped over the surface and would quickly solidify to form the Beilby layer.

This paper describes experiments made to test this. A few measurements were made of the loss of weight which occurs when metals are rubbed. If the metal is removed by tearing off particles, we should expect the soft metals of low tensile strength to lose the most weight. If, however, surface melting or softening under high temperatures is the major factor, we should expect the low melting-point metals to lose the most weight, irrespective of hardness. The results showed the latter to be true.

A more extensive investigation was made of the influence of the relative melting-point of the polisher and the solid. If these high local temperatures really occur at the points of rubbing contact, we should expect relative melting-point to be an important factor in the process. If the polisher melts or softens at a *lower* temperature than the solid, it will melt and flow first and will have comparatively little effect on the solid. Experiment shows that this is the case. Surface flow, polish and the formation of the Beilby layer readily occur on metals, crystals and glasses *provided the melting-point of the polisher is higher than that of the solid*. The relative hardness, as normally measured at room temperature, is comparatively unimportant. The process of polishing is, however, a complicated one, and this statement requires some qualification; it is discussed more fully later in the paper.

EXPERIMENTAL

A simple apparatus was used. The polisher was mounted on an horizontal rotating table 20 cm. in diameter, which could be driven at various speeds.

The test specimen was in the form of a cylinder *ca.* 1.5 cm. in diameter. The bottom end of the cylinder was pressed against the rotating polisher by a ball which fitted into a socket on the top of the specimen. This maintained a normal pressure between the surfaces. Different loads were applied by fitting weights on the arm carrying the ball. The surface of the specimen was ground flat on lead laps impregnated with carborundum and finished on 000 emery paper. The polishing block was usually of camphor, made by pouring the molten camphor into a circular mould. Camphor was chosen because it was soft, had a sharp low melting-point (178°C.), and on melting gave a liquid of low viscosity. The other polishing substances were usually in powder form and were placed on the camphor in which they became embedded.

The surfaces were usually flooded with water during polishing. This does not prevent the occurrence of intense local heating at the points of contact (Bowden and Ridler 1936); its effect is probably to localize the heating and to prevent a large-scale melting of the solids. If no liquid is used, the camphor and the polishing compound may form sticky cakes which adhere to the specimen. Also, in the absence of liquid, the polishing powder may become embedded in the surface of the specimen. The presence of a liquid makes it possible to maintain a greater concentration of the polishing agent under the specimen, and the rate of polishing is greater. In some cases, where the water would lead to chemical attack of the surfaces of the specimen or polisher, it was not used.

Loss of Weight on Sliding

The result of sliding cylinders of lead, Wood's alloy and gallium over a polisher of thick filter paper is shown in Table I. The speed and the load were the same in each case. The loss of weight may be compared with the hardness and the melting-point.

TABLE I—CYLINDER 0.2 CM. DIAM., LOAD 100 G., SPEED 110 CM./SEC.

Metal	Melting-point ° C.	Vickers hardness	Loss of weight in g./cm of sliding $\times 10^{-4}$
Lead	327	5.0	0.6
Wood's alloy	69	25.0	3.7
Gallium	30	6.6	53.0

There is no doubt as to the general trend. Lead is softer than Wood's alloy or gallium yet it loses much *less* weight. It is the low-melting metal which loses the most weight. A microscopic examination of the filter

paper showed small globules of the metal adhering to the fabric of the paper.

The result of rubbing on a block of pure camphor is shown in Table II (camphor m.p. 178° C.).

TABLE II—LOAD 460 G./CM.², SPEED 205 CM./SEC.

Metal	Melting-point ° C.	Vickers hardness	Loss of weight in g./cm. of sliding × 10 ⁻⁷
Lead	327	5.0	0.0
Wood's alloy	69	25.0	3.2
Gallium	30	6.6	165.0

Again the loss in weight depends upon the melting-point and is independent of the hardness of the metal.

Surface Flow and the Relative Melting-point

(i) *Metals on Camphor*—A pin scratch was made in the prepared metal surface, so that the same areas on the surface could be compared in the photomicrographs of various stages. The effect of rubbing Wood's alloy on camphor (m.p. 178° C.) is shown in fig. 1a, b and c, Plate 5, which were taken at different intervals of time. Surface flow readily occurred, the scratches were bridged over and the Beilby layer formed. At normal loads and speeds the surface polish was similar to that produced by polishing metals with a powder. At very high speeds the surface flow becomes a large-scale melting, and globules of melted metal were stuck to cooler areas and smeared out on the camphor.

When tin (m.p. 232° C.) was rubbed on camphor, no sign of surface flow was observed (fig. 2a and b, Plate 5) even on prolonged rubbing at the highest loads and speeds. This cannot be a hardness effect, since Wood's alloy (hardness 25) is much harder than tin (hardness 4). The hardness of each specimen was determined experimentally and is given in Vickers units.

(ii) *Metals on Oxamide*—Oxamide was selected on account of its low melting-point (417° C.) for a powder, and because it is fairly inert chemically. When polishing easily corroded metals it was found that the presence of water caused an etching of the surface by chemical attack due presumably to hydrolysis of the oxamide. For this reason water was not used and the oxamide was added as a powder to the camphor. With glasses and resistant metals water was used.

Tin (m.p. 232° C.), white metal (m.p. 233° C.), type metal (m.p. 275° C.), zinc (m.p. 419° C.) and other metals of which the melting-point was below

or very close to that of oxamide were then tried. In every case surface flow and polish occurred. Fig. 3*a* and *b* (Plate 6) is a typical case and shows characteristic flow. On the other hand, speculum metal (m.p. 745° C.) (see fig. 4*a* and *b*, Plate 6) and copper (m.p. 1083° C.), whose melting-point is above that of oxamide, show no surface flow.

(iii) *Metals on Lead Oxide* (m.p. 888° C.)—Finely powdered lead mon-oxide, made by fusing lead nitrate, was used in the presence of water. Lower melting-point metals such as zinc, white metal, tin, etc. were readily polished by it. Speculum metal (m.p. 745° C., hardness 505), which was untouched by oxamide, now flowed readily. This is shown in fig. 5*a*, *b* and *c*, Plate 7. In (*b*) the smaller scratches have disappeared, and on the different area (*c*), which is nearer the edge of the specimen, the original scratches have practically disappeared. The appearance of many fresh scratches is due to stray emery particles coming out of the lead in which the specimen was embedded. The higher melting metals such as nickel (m.p. 1452° C., hardness 164) and molybdenum (m.p. 2470° C., hardness 234) showed no surface flow. The results for molybdenum are interesting and are shown in fig. 6*a*, *b* and *c*, Plate 7. At first sight (fig. 6*b*) the surface appears greatly changed and the original scratches seem to have disappeared. If, however, the specimen is wiped with caustic soda solution and then examined, it is seen to be quite unchanged, the original scratches and marks are still present. The molybdenum surface had been covered by a thin coating of metallic lead, which could be dissolved off. This lead was probably produced by the reduction of lead oxide by melted camphor, and provides further evidence for the existence of local high temperatures.

(iv) *Metals on Higher Melting Oxides*—High-melting oxides such as chromic oxide (m.p. 1990° C.), stannic oxide (m.p. 1625° C.), ferric oxide (m.p. 1560° C.), zinc oxide (m.p. > 1800° C.), cuprous oxide (m.p. 1235° C.) and magnesium oxide (m.p. 2800° C.) produced surface flow and polish on all the metals tried.

(v) *Glasses and Quartz*—The "softening point" of a glass is necessarily an arbitrary point, since it depends upon the degree of flow required by the experimental method used to measure it. The effect of rubbing lead glass (softening point 489° C.), soda glass (softening point 600° C.), pyrex glass (softening point 815° C.) and quartz (m.p. 1710° C.) on oxamide (m.p. 417° C.) under standard conditions of load and speed is shown in figs. 7, 8, 9 and 10, Plates 8 and 9. The easily fusible lead glass flows readily, soda glass shows a slight flow, with pyrex glass the flow is barely perceptible, and with quartz there is no flow at all.

When high-melting oxides such as chromic oxide, stannic oxide, zinc oxide and ferric oxide were used, they readily caused the flow and polish of quartz and of all the glasses. The effect of stannic oxide is shown in fig. 11, Plate 10. Although the melting-point of cuprous oxide ($1235^{\circ}\text{C}.$) is somewhat below that of quartz ($1710^{\circ}\text{C}.$), it caused surface flow. It is well known, however, that quartz fibres soften in a flame at temperatures well below $1000^{\circ}\text{C}.$

(vi) *Non-metallic Crystals*—Calcite (m.p. $1333^{\circ}\text{C}.$) could not be polished on oxamide (m.p. $417^{\circ}\text{C}.$) or on cuprous oxide (m.p. $1235^{\circ}\text{C}.$), see figs. 12*a*, *b* and *c*, Plate 10. A polisher of higher melting-point, however, zinc oxide (m.p. $> 1800^{\circ}\text{C}.$), gave a good polish in a few minutes (see fig. 13, Plate 11). Results with other non-metallic crystals were not so definite and will be described later.

Surface Flow at Temperatures below the Melting-point

In some cases there was an appreciable flow at temperatures well below the melting-point. This was observed with metals of very high melting-point on polishers which melt above $1000^{\circ}\text{C}.$ For example, nickel (m.p. $1452^{\circ}\text{C}.$), palladium (m.p. $1555^{\circ}\text{C}.$) and molybdenum (m.p. $2470^{\circ}\text{C}.$) were polished by cuprous oxide (m.p. $1235^{\circ}\text{C}.$) and ferric oxide (m.p. $1580^{\circ}\text{C}.$). As will be pointed out later in the discussion, this flow is to be expected, since the resistance of these metals to deforming stresses falls to very low values at temperatures of $1000^{\circ}\text{C}.$, although this may be well below the melting-point. Again, gold (m.p. $1063^{\circ}\text{C}.$) showed some surface flow on oxamide (m.p. $417^{\circ}\text{C}.$), and it was just possible to make molybdenum flow to a slight extent on oxamide if very high loads and speeds were used. Under these conditions, however, the surface had a hammered and beaten appearance which was different from the normal polish.

Similar results were observed when mineralogical crystals were polished on various powders. Crystals of zinc blende (m.p. $1850^{\circ}\text{C}.$), willemite (m.p. $1509^{\circ}\text{C}.$), rutile (m.p. $1640^{\circ}\text{C}.$), galena (m.p. $1114^{\circ}\text{C}.$), rhodonite (m.p. $1215^{\circ}\text{C}.$), corundum (m.p. $2050^{\circ}\text{C}.$), fluorite (m.p. $1360^{\circ}\text{C}.$), beryl (m.p. $1410^{\circ}\text{C}.$) and labradorite (m.p. $1380\text{--}1500^{\circ}\text{C}.$) were mounted in tin and polished on chromic oxide (m.p. $1990^{\circ}\text{C}.$), stannic oxide (m.p. $1625^{\circ}\text{C}.$), ferric oxide (m.p. $1580^{\circ}\text{C}.$) and cuprous oxide (m.p. $1235^{\circ}\text{C}.$). Comparative results were difficult on account of the varied shape of the crystal. It was always the more projecting parts of the crystal which flowed first, because of the higher pressure on these areas.

Again, some flow was observed at temperatures below the melting-point, e.g. zinc blende on cuprous oxide, rutile on ferric oxide, but the results

showed clearly that polishers of lower melting-point than the crystal were very much slower in effect than those melting higher than the crystal. For example, fluorite and beryl flowed rapidly on ferric oxide, but very slowly on cuprous oxide.

DISCUSSION

These experiments show that there is a marked correlation between relative melting-point and surface flow. If the melting-point of the polisher is higher than that of the metal, it rapidly causes surface flow and polish. If the melting-point of the polisher is well below that of the metal, surface flow takes place very slowly or not at all. Camphor (m.p. $178^{\circ}\text{C}.$) readily polishes metals which melt at a lower temperature, e.g. Wood's alloy, but has no influence on tin, lead, white metal and zinc, which melt at a higher temperature. Oxamide (m.p. $417^{\circ}\text{C}.$) will readily cause flow of all these metals, but does not produce any effect on speculum metal or copper which melt at temperatures well above $417^{\circ}\text{C}.$ Lead oxide (m.p. $888^{\circ}\text{C}.$) will polish speculum metal and all metals melting below it, but has little effect on nickel and molybdenum which melt above it. These in turn are readily polished by the high-melting oxides such as chromic oxide, stannic oxide, etc.

Similar results are obtained with glasses, quartz and some non-metallic crystals, calcite, for example, shows no flow on cuprous oxide which melts a little below it, but is readily polished by zinc oxide which melts above it.

It is well known that the mechanical strength of many metals and solids falls to a very low value at temperatures well below the melting-point (Tapsell, 1931; Elam, 1935). The rounding of sharp metal crystals and the low value of tensile strength, hardness, etc. at these temperatures show that metals lose their rigidity and resistance to shear at comparatively low temperatures. For such solids, surface flow would be expected to take place at temperatures well below the melting-point, and experiment shows that, in many case (e.g. gold), this can occur. The rate of flow and polish is however very much less and may take hours, instead of the few minutes required by a high-melting polisher.

The experiments provide strong evidence, not only that high local temperatures occur, but that they play a large part in the process of polishing. In many cases the frictional heat will raise the temperature to a sufficiently high value to cause a real melting of the solid at the points of sliding contact. The molten solid will flow or will be smeared on to cooler areas, and will very quickly solidify to form the "amorphous" Beilby layer. Polishing under these conditions is rapid. If the sliding is gentle or the

melting-point of the polisher is low, the surface of the solid may not reach the temperature of melting. Polish and surface flow may still occur under these conditions, provided the temperature reaches a point at which the mechanical strength of the solid is sufficiently low for it to yield under the applied stress or under surface tension forces. Polishing under these conditions is a slow process.

The relative hardness of solid and polisher as normally measured at room temperature is comparatively unimportant. This is shown clearly in the case of Wood's alloy and tin on camphor, or speculum metal and nickel on lead oxide. The harder metal of low melting-point is polished, while the softer metal of higher melting-point hardly flows at all. Similarly zinc oxide which is comparatively soft (Mohs' hardness 4) readily polishes quartz (Mohs' hardness 7). The amount of surface flow is governed, not by the properties of the solids at room temperature, but by their relative mechanical properties *at the high temperature of the sliding surfaces*.

A simple calculation shows that the amount of heat liberated during sliding is sufficient to raise the temperature to the melting-point, and to melt a considerable quantity of the solid. If the coefficient of friction is 0.2, the load is 500 g./cm.² and the sliding speed 50 cm./sec., the amount of heat Q liberated at the interface is given by

$$Q = \frac{0.2 \cdot 500 \cdot 981 \cdot 50}{4 \cdot 18 \cdot 10^7} = 0.117 \text{ cal./sec.}$$

If the polishing is carried on for (say) 10 min., the amount of heat liberated is 70 cal.

The amount of heat Q required to raise a monolayer of copper from room temperature to the melting-point, and to melt it, is

$$Q = 2.0 \cdot 10^{-7} (0.093 \times 1063 + 42) = 2.8 \cdot 10^{-5} \text{ cal./sq. cm.,}$$

since $2.0 \cdot 10^{-7}$ g. is the mass per cm.² of a monolayer of copper, 0.093 is the specific heat and 42 is the latent heat of fusion. So that polishing for 10 min. will liberate enough heat to raise to the melting-point and to melt approximately 10^6 monolayers of copper. Electron diffraction experiments show that the thickness of the Beilby layer is approximately 10 monolayers. A great part of the heat generated is, of course, dissipated by conduction and radiation, but it is clear that the retention of a very small fraction of it (one part in 100,000) would be sufficient to form, by melting, the Beilby layer. It should be emphasized that this heat is not liberated uniformly over the whole square centimetre of surface, but is concentrated at the points of sliding contact so that the probability of high local temperatures and melting is very great.

The Structure of the Beilby Layer and the Mechanism of its Formation

The surface of a polished metal gives an electron diffraction pattern of two diffuse rings (R. C. French 1933; Raether 1933), while non-metallic crystals give, in some cases, a similar pattern and in others a blurred spot pattern (Jenkins 1934; Hopkins 1936; Finch 1936). There is general agreement that the spot pattern is caused by an orientated layer of micro-crystals, but there is some difference of opinion as to the correct interpretation of the diffuse ring pattern. R. C. French (1933), Darbyshire and Dixit (1933), Thomson (1934), and Finch and Quarrell (1936) accept this as evidence that the polished layer is truly amorphous, while Germer (1936) and Kirchner (1932, 1937) consider that the diffuse pattern is merely due to levelling, by polishing, of the surface projections, and that the surface is in reality micro-crystalline.

The distinction between the micro-crystalline and the amorphous state may, in the limit, be an academic one. This paper suggests that an important part of the polishing process is an intense local heating of the solid to its softening or melting-point, followed by a very rapid cooling. Since polishing is often done in the presence of water or other liquids the cooling may be very rapid indeed. Such treatment would greatly reduce the crystal size, and with most metals would produce a very finely crystalline or "vitreous-like" layer. The precise extent to which the crystalline structure is destroyed will naturally depend on the conditions of polishing, and on the physical properties of the solid. If the solid is capable of forming a glass when it is solidified by rapid cooling the Beilby layer may be amorphous when first formed, but it will not necessarily remain in this state. Experiments with thin films of sputtered metal show that small crystal aggregates grow rapidly at room temperature.

If the polishing is carried out in air the local heating may cause rapid surface oxidation and we should expect the surface to consist of oxide. Evidence that this is so is provided by the work of Dobinski (1936) and Preston and Bircumshaw (1936) who found that the diffraction pattern given by copper and aluminium was characteristic of the metallic oxide.*

It is probable that the Beilby layer produced under most conditions of

* A more recent paper by Dobinski (1937) confirmed these observations for aluminium and copper and also showed that the polished layer on iron, nickel, cadmium and tin was metallic oxide. No oxide, however, was observed on gold and silver. This provides further evidence for the suggestion put forward in this paper that local high temperatures occur, since the oxides on these two metals decompose at 205° C. and 180° C. respectively, whereas the other metallic oxides are stable at high temperatures.

polishing will not be homogeneous, but will consist of rapidly cooled solid containing very small aggregates and micro-crystals of varying size, in which oxide and even particles of the polishing powder may be embedded.

These local high temperatures may also play an important part in the so-called "cold" working of metals. When the metal is worked, slip occurs in the crystals. The deformation proceeds in jerks (Joffé 1928, p. 50), so that the actual rate of sliding may be high. Since the pressure and friction at the points of contact may be considerable, the temperature at these points may momentarily be raised to a sufficiently high value to cause softening or melting with consequent flow. The result of this will be to produce smaller crystals which are cemented together by "amorphous" metal. Farren and Taylor (1925) have shown that from 5 to 15% of the work of deformation is retained by the metal, the balance being transformed into heat, and Stepanow's (1933) measurements of the conductivity of crystals undergoing distortion provide some evidence that such local high temperatures may occur.

In a similar way the failure of metals in fatigue tests may under some conditions be due to high-temperature oxidation at the slip planes by the air included in the metal.

The Wear and Corrosion of Sliding Surfaces

The fact that these high local temperatures are reached when one surface rubs on another must have an important bearing on the process of wear of sliding surfaces. This is borne out by the loss of weight experiments, which suggest that wear is determined more by the mechanical properties at the high temperature of sliding than by the properties at room temperature. Under some conditions of sliding, wear must be due to a local melting away of the surface irregularities. Even in the presence of a good lubricant these local-high temperatures may still occur.

The intense local heating at the points of contact will also cause oxidation and corrosion of the sliding surfaces. Several examples of such chemical action were observed during these experiments:

(i) The etching action of oxamide is probably due to the local heating of the aqueous solution, which is known to hydrolyse to oxalic acid on boiling.

(ii) When copper was polished on powdered anhydrous sodium carbonate the surface was attacked rapidly, and green spots were formed. These were probably a basic copper carbonate, formed by the local heating of the two reactants in close contact.

(iii) The reduction of lead oxide to lead by camphor. The lead formed on

the grains must be rubbed or melted off on to the metal surface of the specimen, where it forms a very thin coating.

(iv) When mercuric chloride or mercuric oxide were used as polishing powders, they rapidly decomposed to give a polished surface of the metal amalgam. The temperature needed to bring this about is only 100° C. and so is easily attained.

(v) If tin is slid on glass there is a marked discoloration and corrosion at the points of contact, which is due to local heating and oxidation of the tin.

Whenever an oxidizable metal or solid is rubbed in air, it is probable that the local high temperatures at the points of contact will cause rapid surface oxidation. We should expect, for example, to find oxide on the surface of a metal which has been freshly ground with emery paper, and Evans (1927) has shown that this is the case. Many of the observations on the oxidation and corrosion of sliding surfaces may be attributed to heating at the points of contact, rather than to the action of air at high pressure trapped in the surface cracks, as has been suggested (see Desch 1934, p. 104).

We are greatly indebted to Professor R. S. Hutton and Dr D. Stockdale of the Metallurgical Laboratory for valuable assistance, to the Chemical Society and the Department of Scientific and Industrial Research for grants for apparatus, and to Gonville and Caius College for a studentship granted to one of us (T. P. H.).

SUMMARY

The process of polishing is greatly influenced by the relative melting-point of the polisher and the solid. The relative hardness is comparatively unimportant. Experiment suggests that surface flow is brought about by an intense local heating of the surface irregularities to the melting or softening point. The molten or softened solid flows or is smeared over the surface, and very quickly solidifies to form the polished Beilby layer.

These local high temperatures also play an important part in the wear and corrosion of sliding surfaces.

REFERENCES

- Adam, N. K. 1927 *Nature, Lond.*, **119**, 162, 279-80.
Beilby, Sir George 1921 "Aggregation and Flow of Solids," 1sted. London: Macmillan and Co.
Bowden, F. P. and Rüdler, K. E. W. 1935 *Proc. Camb. Phil. Soc.* **31**, 431-2.

- Bowden, F. P. and Ridler, K. E. W. 1936 *Proc. Roy. Soc. A*, **154**, 640-56.
 Darbyshire, J. A. and Dixit, K. R. 1933 *Phil. Mag.* **16**, 961-74.
 Desch, C. H. 1934 "The Chemistry of Solids," 1st ed. New York: Cornell Univ. Press.
 Dobinski, S. 1936 *Nature, Lond.*, **138**, 31.
 — 1937 *Phil. Mag.* **23**, 397-408.
 Elam, C. F. 1935 "The Distortion of Metal Crystals," 1st ed. Oxford: Univ. Press.
 Evans, U. R. 1927 *J. Chem. Soc.* **2**, 1020-40.
 Farren, W. S. and Taylor, G. I. 1925 *Proc. Roy. Soc. A*, **107**, 422-51.
 Finch, G. I. 1936 *Nature, Lond.*, **138**, 1010.
 Finch, G. I. and Quarrell, A. G. 1936 *Nature, Lond.*, **137**, 516-19.
 French, J. W. 1927 *Nature, Lond.*, **119**, 527.
 French, R. C. 1933 *Proc. Roy. Soc. A*, **140**, 637-52.
 Germer, L. H. 1936 *Phys. Rev.* **49**, 163-6.
 Hamburger, L. 1932 *Nature, Lond.*, **130**, 435-6.
 Herschel, J. F. W. 1830 "Encyclopaedia Metropolitana," "Light".
 Hopkins, H. G. 1936 *Phil. Mag.* **21**, 820-9.
 Jenkins, R. O. 1934 *Phil. Mag.* **17**, 457-66.
 Joffé, A. F. 1928 "The Physics of Crystals," 1st ed. New York: McGraw-Hill Co.
 Kirchner, F. 1932 *Nature, Lond.*, **129**, 545.
 — 1937 *Ann. Phys., Lpz.* **28**, 21-7.
 Macaulay, J. M. 1926 *Nature, Lond.*, **116**, 339.
 — 1927 *Nature, Lond.*, **119**, 13.
 — 1929-1932 *J.R. Tech. Coll. Glasg.* **2**, 378-85.
 Newton, Sir Isaac 1730 "Opticks," 1931 (reprint from 4th ed.). London: Bell and Sons.
 Preston, G. D. and Bircumshaw, L. L. 1936 *Phil. Mag.* **23**, 654-64.
 Raether, H. 1933 *Z. Phys.* **86**, 82-104.
 Rayleigh, Lord 1901a *Proc. Roy. Instn.* **16**, 563-70.
 — 1901b *Nature, Lond.*, **64**, 385-8.
 Stepanow, A. W. 1933 *Z. Phys.* **81**, 560-4.
 Tapsell, H. J. 1931 "Creep of Metals," 1st ed. London: Oxford Univ. Press.
 Thomson, G. P. 1934 *Phil. Mag.* **18**, 640-56.

DESCRIPTION OF PLATES

Plate 5

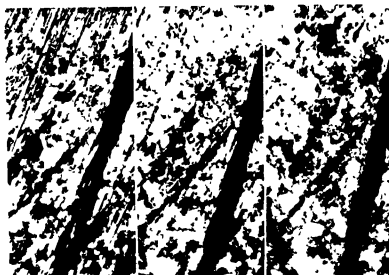
FIG. 1—Wood's alloy (m.p. 69° C.) on camphor (m.p. 178° C.). Load 500 g./cm.² Speed 104 cm./sec. No liquid medium. $\times 220$. Surface flow.

FIG. 2—Tin (m.p. 232° C.) on camphor (m.p. 178° C.). Load 460 g./cm.² Speed 205 cm./sec. No liquid medium. $\times 235$. No surface flow.

Plate 6

FIG. 3—Zinc (m.p. 419° C.) on oxamide (m.p. 417° C.). Load 617 g./cm.² Speed 104 cm./sec. No liquid medium. $\times 212$. Surface flow.

FIG. 4—Speculum metal (m.p. 745° C.) on oxamide (m.p. 417° C.). Load 1000 g./cm.² Speed 205 cm./sec. No liquid medium. $\times 232$. No surface flow.

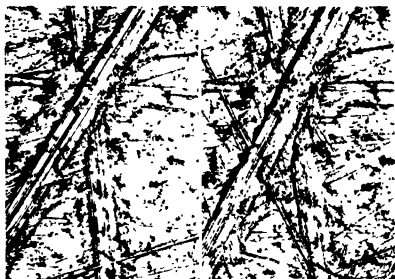


a. Start

b. 1 hour

c. 2 hours

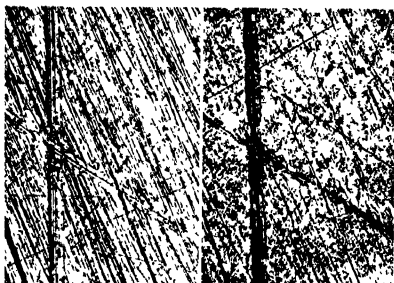
FIG. 1



a. Start.

b. 2½ hours

FIG. 2



a. Start

b. 1/2 hour

FIG. 3



a. Start.

b. 1/2 hour

FIG. 4

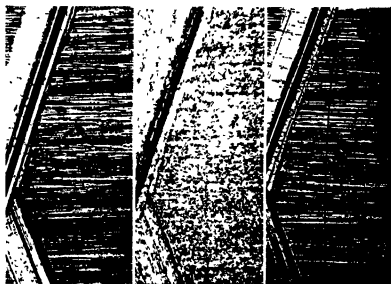


a Start

b 1 hour

c 1 hour (another area).

FIG. 5

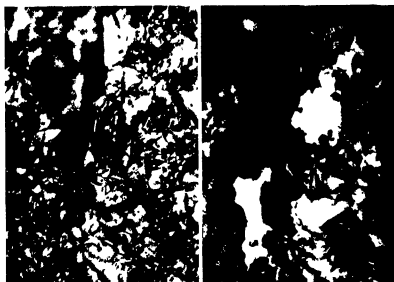


a Start

b 1 hour (lead deposit)

c 1 1/2 hours (lead removed)

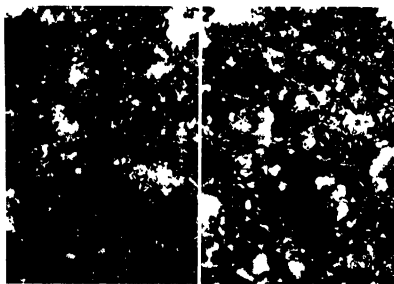
FIG. 6



a Start

b $\frac{1}{2}$ hour

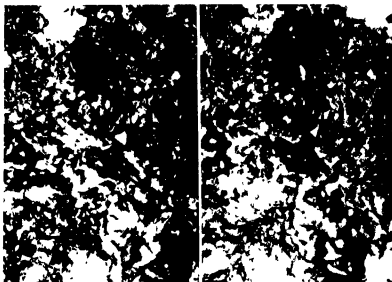
FIG. 7



a Start.

b, 1 hour

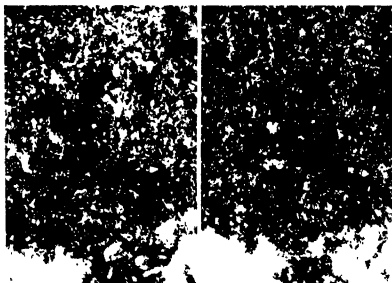
FIG. 8



a Start

b 1 hour

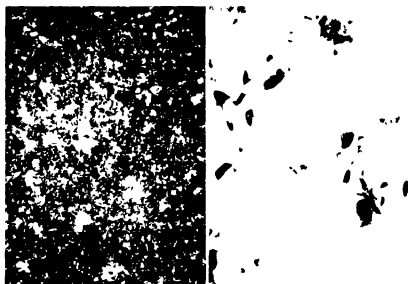
FIG. 9



a Start.

b 1 hour

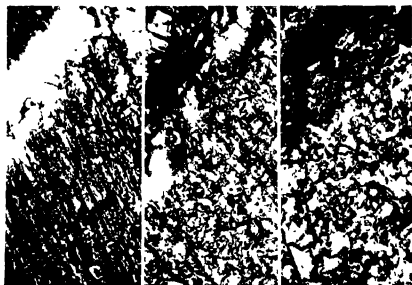
FIG. 10



a Start

b $\frac{1}{2}$ hour

FIG. 11

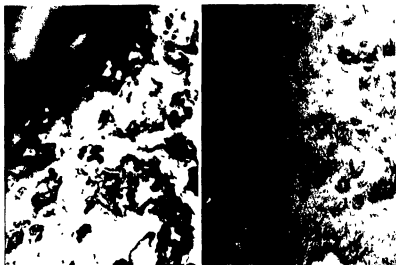


a Start

b $\frac{1}{2}$ hour (oxamide)

c $\frac{1}{2}$ hour (cuprous
oxide)

FIG. 12



a 5 mm

b Same, another area

FIG. 13

Plate 7

FIG. 5—Speculum metal (m.p. $745^{\circ}\text{C}.$) on lead oxide (m.p. $888^{\circ}\text{C}.$). Load 1000 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. Surface flow.

FIG. 6—Molybdenum (m.p. $2470^{\circ}\text{C}.$) on lead oxide (m.p. $888^{\circ}\text{C}.$). Load 334 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 278$. No surface flow.

Plate 8

FIG. 7—Lead glass (s.p. $469^{\circ}\text{C}.$) on oxamide (m.p. $417^{\circ}\text{C}.$). Load 493 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. Surface flow.

FIG. 8—Soda glass (s.p. $600^{\circ}\text{C}.$) on oxamide (m.p. $417^{\circ}\text{C}.$). Load 378 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. Slight surface flow.

Plate 9

FIG. 9—Pyrex glass (s.p. $815^{\circ}\text{C}.$) on oxamide (m.p. $417^{\circ}\text{C}.$). Load 327 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. No surface flow.

FIG. 10—Quartz (m.p. $1710^{\circ}\text{C}.$) on oxamide (m.p. $417^{\circ}\text{C}.$). Load 327 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. No surface flow.

Plate 10

FIG. 11—Quartz (m.p. $1710^{\circ}\text{C}.$) on stannic oxide (m.p. $1625^{\circ}\text{C}.$). Load 327 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. Surface flow.

FIG. 12—Calcite (m.p. $1333^{\circ}\text{C}.$) on oxamide (m.p. $417^{\circ}\text{C}.$) and cuprous oxide (m.p. $1235^{\circ}\text{C}.$). Load 350 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. No surface flow.

Plate 11

FIG. 13—Calcite (m.p. $1333^{\circ}\text{C}.$) on zinc oxide (m.p. $1800^{\circ}\text{C}.$). Load 350 g./cm.^2 Speed 52 cm./sec. Aqueous medium. $\times 330$. Surface flow.

A Generalization of the Equations of the Self-Consistent Field for Two-Electron Configurations

BY A. F. STEVENSON, M.A., Ph.D., *Department of Applied Mathematics, University of Toronto*

(Communicated by D. R. Hartree, F.R.S.—Received 13 February 1937)

1—INTRODUCTION

The most successful general method so far devised for dealing with many-electron atoms is that of the self-consistent field (abbreviated in what follows to "s.c.f."). If greater accuracy is required than is obtainable with the method as ordinarily used (either with or without exchange), either the so-called "configuration interaction" must be taken into account—usually a very laborious procedure—or else more complicated (variational) methods must be used, which must be designed separately for each particular case, and in which the concept of each electron being assigned to its own "orbit" is usually abandoned.

It would seem desirable, therefore, to have, if possible, some *general* method which will increase the accuracy of the calculations without taking into account configuration interaction, and which will still allow the conceptual features of the s.c.f. method (i.e. the assignment of "orbits") to be retained. In this paper such a method is developed for the case of two-electron configurations in Russell-Saunders coupling. The method consists in assuming a form for the wave function which is similar to that used in the s.c.f. method, except that the proper spatial symmetry is allowed for (which is not so in the case of the s.c.f. equations *without* exchange), and further, an adjustable function of Θ , the angle between the radii vectores to the two electrons, is inserted as a multiplying factor. The usual variational method is then applied, and yields differential equations for the two radial functions which are similar to those of the ordinary theory, together with an equation for the angular function

The equations are given both for the case when exchange degeneracy is neglected and when it is included; in the latter case, the equations prove to be quite complicated, though a numerical solution would be possible. For the equations without exchange, however, the increase in complexity over the usual equations of the s.c.f. is not very great, and where solutions of the latter are available, those of the corresponding generalized equations

should be fairly readily obtainable. It is well known that the inclusion of θ in some manner (or equivalently the inter-electronic distance) has materially increased the accuracy of the calculated energy values.[†] On the whole, it seems probable that the present method *without* exchange should yield better results for the energy than does the s.c.f. *with* exchange.

For the most part, the two valence electrons are treated as moving in a *given* central (not necessarily Coulomb) field and under the influence of their mutual repulsion, so that the polarization of the atom core is neglected; but the case where the core electrons are taken into account is also briefly considered.

2—THE FORM ASSUMED FOR THE WAVE FUNCTION

We consider the two electrons, labelled 1 and 2, as being in orbits with principal and azimuthal quantum numbers n_1, l_1 and n_2, l_2 , combining to form a state specified by the quantum number L , where $|l_1 - l_2| \leq L \leq l_1 + l_2$, and choose as co-ordinates defining the system: α, β, γ Eulerian angles defining the position of the triangle consisting of the nucleus and the two electrons (α, β being the angular polar co-ordinates of electron 1 relative to the nucleus), r_1, r_2 the distances of the electrons from the nucleus, and θ the angle between \mathbf{r}_1 and \mathbf{r}_2 . We consider the two electrons as moving under the influence of their mutual repulsion and of a given central field. In atomic units the potential energy of the system is then

$$V = -V(r_1) - V(r_2) + 1/r_{12}, \quad (1)$$

where $V(r)$ is the potential of the central field.

The Hamiltonian for this system of co-ordinates has been given by Hylleraas (1928). In atomic units (the energy being in units of the ionization energy of the H-atom, and being *negative*) it is

$$\begin{aligned} H\psi = & -\frac{1}{r_1^2} \frac{\partial}{\partial r_1} \left(r_1^2 \frac{\partial \psi}{\partial r_1} \right) - \frac{1}{r_2^2} \frac{\partial}{\partial r_2} \left(r_2^2 \frac{\partial \psi}{\partial r_2} \right) + \frac{1}{r_1^2} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + A_1 \psi \right] \\ & + \frac{1}{r_2^2} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + A_2 \psi \right] + 2V\psi, \quad (2) \end{aligned}$$

[†] The best-known example being the calculations of Hylleraas (1928, 1929) on the ground state of He. The form adopted by Hylleraas is quite different from that used here. Although the present method probably would not yield a result for the energy comparable to that of Hylleraas for this particular case, it has the advantage that it is applicable to *any* two-electron configuration.

where V is given by (1), and

$$\left. \begin{aligned} A_1\psi &= (\cot^2\Theta + 2\cot\Theta \cot\beta \cos\gamma + \cot^2\beta) \frac{\partial^2\psi}{\partial\gamma^2} + \frac{1}{\sin\beta} \frac{\partial}{\partial\beta} \left(\sin\beta \frac{\partial\psi}{\partial\beta} \right) \\ &\quad + \frac{1}{\sin^2\beta} \frac{\partial^2\psi}{\partial\alpha^2} + 2\cot\beta \sin\gamma \frac{\partial^2\psi}{\partial\Theta \partial\gamma} - 2\cos\gamma \frac{\partial^2\psi}{\partial\Theta \partial\beta} - 2\frac{\sin\gamma}{\sin\beta} \frac{\partial^2\psi}{\partial\Theta \partial\alpha} \\ &\quad + 2\cot\Theta \sin\gamma \frac{\partial^2\psi}{\partial\beta \partial\gamma} - \frac{2}{\sin\beta} (\cot\beta + \cot\Theta \cos\gamma) \frac{\partial^2\psi}{\partial\alpha \partial\gamma}, \\ A_2\psi &= \frac{1}{\sin^2\Theta} \frac{\partial^2\psi}{\partial\gamma^2}. \end{aligned} \right\} \quad (3)$$

The element of volume is (Breit 1930)

$$dv = r_1^2 r_2^2 \sin\Theta dr_1 dr_2 d\Theta dv,$$

where $dv = \sin\beta d\alpha d\beta d\gamma$ is the element of volume of the space of the Eulerian angles. The range of variation of the angular co-ordinates is 0 to 2π for α and γ , 0 to π for β and Θ .

Before considering the general case, we may illustrate the principle of the method by considering the simplest possible configuration, namely s^2 . The resulting state, 1S , is then spherically symmetrical, so that the wave function is *accurately* of the form

$$\psi = \psi(r_1, r_2, \Theta). \quad (4)$$

The s.c.f. approximation for this case consists in assuming

$$\psi = R(r_1) R(r_2) \quad (5)$$

and disregards the dependence on Θ altogether. An evident improvement on (5) is to assume

$$\psi = R(r_1) R(r_2) X(\Theta), \quad (6)$$

which will allow to some extent for the dependence on Θ . The general case is more complicated in that the s.c.f. solution itself depends on Θ (in addition to α, β, γ), but we can again improve this by inserting an additional function of Θ as a factor.

Proceeding now to the general case, we must first find the form of the wave function when the assumption corresponding to the s.c.f. is made, but allowing for the proper spatial symmetry of the problem. This can be deduced from group-theoretical considerations. Disregarding exchange for the moment, it is known that, associated with the L state, there is a set of

independent wave-functions $\psi_M^L (M = -L, \dots, L)$ which are accurately of the form†

$$\psi_M^L = \sum_{M'=-L}^L e^{-i(M\alpha+M'\gamma)} y_{MM'}^L(\beta) G_{M'}(r_1, r_2, \Theta), \quad (7)$$

where

$$y_{MM'}^L(\beta) = \sum_k (-1)^k \frac{[(L+M)! (L-M)! (L+M')! (L-M')!]^{\frac{1}{2}}}{(L-M-k)! (L+M'-k)! k! (k+M-M')!} \\ \times (\cos \beta/2)^{2L+M'-M-2k} (\sin \beta/2)^{2k+M-M'}, \quad (8)$$

the summation with respect to k in (8) being from $\max(0, M'-M)$ to $\min(L-M, L+M')$. The functions of α, β, γ occurring on the right-hand side of (7) form a matrix which furnishes a $(2L+1)$ -dimensional irreducible representation of the rotation group. On the other hand, the s.c.f. approximation (without exchange) assumes that ψ is of the form

$$R_1(r_1) R_2(r_2) f(\alpha, \beta, \gamma, \Theta),$$

and the zero approximations to the wave functions are linear combinations of such terms. Comparing with (7), we see that the s.c.f. assumption, generalized to allow for spatial symmetry, but not including exchange, is of the form

$$\psi_M^L = R_1(r_1) R_2(r_2) F_M(\alpha, \beta, \gamma, \Theta), \quad (9)$$

where

$$F_M = \sum_{M'=-L}^L e^{-i(M\alpha+M'\gamma)} y_{MM'}^L(\beta) g_{M'}(\Theta), \quad (10)$$

the functions R_1, R_2 being independent of M and M' , and the functions $g_{M'}$ being independent of M . The explicit form of $g_{M'}$ can be deduced from the fact that (9) must be accurately a solution of the wave equation when the interaction $1/r_{12}$ is put equal to zero ‡. In this case the radial functions R must satisfy

$$\left. \begin{aligned} \frac{1}{r_1^2} \frac{d}{dr_1} \left(r_1^2 \frac{dR_1}{dr_1} \right) + \left[E^{(1)} + 2V(r_1) - \frac{l_1(l_1+1)}{r_1^2} \right] R_1 &= 0, \\ \frac{1}{r_2^2} \frac{d}{dr_2} \left(r_2^2 \frac{dR_2}{dr_2} \right) + \left[E^{(2)} + 2V(r_2) - \frac{l_2(l_2+1)}{r_2^2} \right] R_2 &= 0, \end{aligned} \right\} \quad (11)$$

† Wigner (1931, p. 233, formula (18), and p. 180, formula (27)). We have changed Wigner's notation slightly. Cf. also Breit (1930).

‡ An alternative procedure would be to use the expression (21) below for F_M in terms of polar co-ordinates, and then to express this in terms of the new co-ordinates $\alpha, \beta, \gamma, \Theta$; but this seems to be difficult to carry out.

the total energy being $E^{(1)} + E^{(2)}$. Hence, substituting (9) in the wave equation, and using (11) and the form (2) for the Hamiltonian with $1/r_{12}$ put equal to zero, we find that F_M must satisfy the two equations

$$\frac{1}{\sin \Theta} \frac{\partial}{\partial \Theta} \left(\sin \Theta \frac{\partial F_M}{\partial \Theta} \right) + A_1 F_M + l_1(l_1 + 1) F_M = 0, \quad (12)$$

$$\frac{1}{\sin \Theta} \frac{\partial}{\partial \Theta} \left(\sin \Theta \frac{\partial F_M}{\partial \Theta} \right) + A_2 F_M + l_2(l_2 + 1) F_M = 0. \quad (13)$$

Using (10), (13) gives

$$\frac{1}{\sin \Theta} \frac{d}{d\Theta} \left(\sin \Theta \frac{dg_{M'}}{d\Theta} \right) + \left[l_2(l_2 + 1) - \frac{M'^2}{\sin^2 \Theta} \right] g_{M'} = 0, \quad (14)$$

so that

$$g_{M'}(\Theta) = c_{M'}(L, l_1, l_2) P_{l_2}^{M'}(\cos \Theta), \quad (15)$$

where the constants $c_{M'}$ may depend on L, l_1, l_2 in addition to M' , but must be independent of M . To find the constants $c_{M'}$ (or rather their ratio, which is, of course, all that can be determined), we must use (12). We easily find, in fact, that in order that (12) may be satisfied identically in α and γ , the following relation must be satisfied identically in β and Θ for all values of M' ,† in which we write for simplicity $y_{M'}$ in place of $y_{M'M'}^{L}(\beta)$:

$$\begin{aligned} & [l_1(l_1 + 1) - l_2(l_2 + 1) - L(L + 1) + 2M'^2] y_{M'} g_{M'} \\ & - (M' + 1)^2 \cot \Theta \cot \beta y_{M'+1} g_{M'+1} - (M' + 1) \cot \beta y_{M'+1} \frac{dg_{M'+1}}{d\Theta} \\ & - (M' + 1) \cot \Theta \frac{dy_{M'+1}}{d\beta} g_{M'+1} - \frac{dy_{M'+1}}{d\beta} \frac{dg_{M'+1}}{d\Theta} \\ & - (M' - 1)^2 \cot \Theta \cot \beta y_{M'-1} g_{M'-1} + (M' - 1) \cot \beta y_{M'-1} \frac{dg_{M'-1}}{d\Theta} \\ & + (M' - 1) \cot \Theta \frac{dy_{M'-1}}{d\beta} g_{M'-1} - \frac{dy_{M'-1}}{d\beta} \frac{dg_{M'-1}}{d\Theta} = 0. \quad (16) \end{aligned}$$

† Use has been made of the fact that $g_{M'}$ satisfies (14), and that $y_{M'}$ satisfies the equation

$$\frac{1}{\sin \beta} \frac{d}{d\beta} \left(\sin \beta \frac{dy_{M'}}{d\beta} \right) + \left[-\frac{M^2 + M'^2}{\sin^2 \beta} + 2MM' \frac{\cos \beta}{\sin^3 \beta} + L(L + 1) \right] y_{M'} = 0,$$

which can be deduced, for instance, from the fact that $e^{i(M\alpha + M'\gamma)} y_{M'M'}^{L}(\beta)$ is a solution of equation (14) of Breit's paper (1930), and taking into account the difference in notation. We have further simplified by taking $M = 0$, which is permissible since the results must be independent of M . (16) holds for any value of M' , if it be remembered that $P_b^a(x) = 0$ if $a > b$.

For different values of M' , (16) gives a series of relations to determine the ratios of the c 's. We may further use the fact that, on account of the difference in behaviour of odd and even terms with respect to reflexion in the origin (Wigner 1931, p. 234),

$$G_{-M'}(r_1, r_2, \Theta) = (-1)^{L+l_1+l_2+M'} G_{M'}(r_1, r_2, \Theta),$$

so that

$$c_{-M'} = (-1)^{L+l_1+l_2+M'} c_{M'}. \quad (17)$$

The general case seems to be rather difficult to handle, but for particular values of L the results are easily obtained. On account of (17), we must distinguish between odd and even terms. For S, P, D terms, the results are as follows:

S: $L = 0$, $l_1 = l_2$. The only coefficient is c_0 .

Even P: $L = 1$, $l_1 = l_2$. $c_0 = 0$, $c_{-1} = c_1$

Odd P: $L = 1$, $|l_1 - l_2| = 1$. $c_{-1} = -c_1$, $c_1/c_0 = -2^{\frac{1}{2}}/f(l_1, l_2)$.

Even D: $L = 2$, $|l_1 - l_2| = 0, 2$. $c_{-1} = -c_1$, $c_{-2} = c_2$,

$$c_1/c_0 = (6^{\frac{1}{2}}/12) [f(l_1, l_2) - 6]/l_2(l_2 + 1),$$

$$c_2/c_0 = (2/3) 6^{\frac{1}{2}} [f(l_1, l_2) - 6]/l_2(l_2 + 1) [f(l_1, l_2) + 2].$$

Odd D: $L = 2$, $|l_1 - l_2| = 1$. $c_0 = 0$, $c_{-1} = c_1$, $c_{-2} = -c_2$,

$$c_2/c_1 = [4 - f(l_1, l_2)]/2(l_2 - 1)(l_2 + 2) \dagger$$

$f(l_1, l_2)$ denotes $(l_1 - l_2)(l_1 + l_2 + 1)$.

We can now write (9) in the form

$$\psi_M^L(1, 2) = \frac{1}{r_1} P_1(r_1) \frac{1}{r_2} P_2(r_2) F_M(\alpha, \beta, \gamma, x), \quad (18)$$

$$\text{where} \quad F_M = e^{-iM\alpha} \sum_{M'=-L}^L c_{M'} e^{-iM'\gamma} y_{MM'}^L(\beta) P_h^{M'}(x). \quad (19)$$

We have written x for $\cos \Theta$ ($-1 \leq x \leq 1$), and have followed Hartree in putting $R(r) = P(r)/r$. The summation in (19) need, of course, only be extended from $-l_2$ to l_2 if $L > l_2$.[†]

We now generalize the assumption (18), in a way corresponding to (6), to

$$\psi_M^L(1, 2) = X(x) \frac{1}{r_1} P_1(r_1) \frac{1}{r_2} P_2(r_2) F_M(\alpha, \beta, \gamma, x), \quad (20)$$

[†] $l_2 = 1$ is a possibility, but in that case there is no coefficient c_2 .

[‡] For analytical simplicity, electron 2 would be chosen to be that with the least azimuthal quantum number; but the results are unaltered by interchanging the electrons.

where F_M is again given by (19), and $X(x)$ is an adjustable function to be determined by the variational method. Since (20) is more general than (18), and includes (18) as a special case, it must necessarily give a more accurate result for the energy. From (19) and (17), we see that in the case of S terms or even P terms, or in the case $l_2 = 1$, $L = l_1$, F_M is equal to the product of a function of x and a function of α, β, γ , and in this case, as we shall see, it may be convenient to absorb the " x " part of F_M into the function X in (20). If, as will be shown, the functions P_1, P_2, X are independent of M , the assumption (20) retains the proper spatial symmetry. Further, it will appear that X is unaltered by interchange of the electrons, as should evidently be the case. A more general assumption than (20) would be to assume that each separate G_M in (7) is of the form $R_1^{(M)}(r_1) R_2^{(M)}(r_2) X^{(M)}(x)$, but this would lead to equations too complicated to handle, and further would destroy the simple analogy with the s.c.f. theory. The assumption (20) appears to be the most general one which preserves this analogy and has the proper spatial symmetry.

We shall also require the result of interchanging the electrons in (20). This is most simply obtained by expressing F_M in terms of the polar co-ordinates of the two electrons. We have, in fact,[†]

$$F_M(1, 2) = e^{iM\phi_2} \sum_{\mu} N_{\mu, l_1} N_{M-\mu, l_2} s_{L, \mu, M-\mu}^{(l_1, l_2)} e^{i\mu(\phi_1 - \phi_2)} P_{l_1}^{\mu}(\cos \theta_1) P_{l_2}^{M-\mu}(\cos \theta_2), \quad (21)$$

where $s_{L, \mu, M-\mu}^{(l_1, l_2)}$ are the coefficients in the "Clebsch-Gordan series" (Weyl 1931, p. 128) and the N 's are normalizing factors for the surface harmonics. Interchanging the co-ordinates of electrons 1 and 2, and then changing the variable of summation from μ to $\mu' = M - \mu$, and using the fact that (Wigner 1931, p. 206, footnote) $s_{L, \mu, \nu}^{(l_1, l_2)} = (-1)^{L+l_1+l_2} s_{L, \nu, \mu}^{(l_2, l_1)}$, we see that the result is to multiply (21) by the factor $(-1)^{L+l_1+l_2}$ and interchange l_1 and l_2 . Thus, corresponding to (20), we have (assuming that X is unaltered by the interchange)

$$\psi_M^L(2, 1) = (-1)^{L+l_1+l_2} X(x) \frac{1}{r_1} P_2(r_1) \frac{1}{r_2} P_1(r_2) \bar{F}_M(\alpha, \beta, \gamma, x), \quad (22)$$

where the bar indicates—here and in what follows—that l_1 and l_2 are to be interchanged—i.e. \bar{F}_M is given by the right-hand side of (19) with $\bar{c}_M = c_M(L, l_2, l_1)$ in place of c_M . Interchanging the co-ordinates of the electrons is thus equivalent to interchanging their quantum numbers to within a \pm sign.

[†] This follows from Wigner (1931, p. 206, formula (18a)) on omitting the radial factors of the wave functions.

3—THE DIFFERENTIAL EQUATIONS, WITHOUT EXCHANGE

Using (20), we now apply the variational principle in the form used by Fock (1930),

$$\int \delta \psi^* (H - E) \psi d\tau = 0. \quad (23)$$

The analysis can be considerably simplified by using (12) and (13), and by noting that consideration of the complicated operator A_1 in the Hamiltonian can be avoided, and attention confined to the much simpler operator A_2 . For since H is unaltered by interchanging the two electrons, an inspection of (2) shows that if a transformation of the Eulerian angles α, β, γ which corresponds to interchange of the electrons be made,† the operators A_1, A_2 are simply interchanged. If we denote the original variables by (α, β, γ) and the new variables by $(\alpha', \beta', \gamma')$, the explicit form of the transformation referred to is, from elementary spherical trigonometry,

$$\cos \beta' = \cos \beta \cos \Theta - \sin \beta \sin \Theta \cos \gamma,$$

$$\sin(\alpha' - \alpha) \sin \beta' = \sin \Theta \sin \gamma,$$

$$\sin \gamma' \sin \beta' = -\sin \beta \sin \gamma.$$

We shall refer to this transformation as the transformation T , and its inverse as T^{-1} . From the above explicit form it is easily shown that $\partial(\alpha, \beta, \gamma)/\partial(\alpha', \beta', \gamma') = \sin \beta'/\sin \beta$, so that, as we might expect, the element of volume $\sin \beta d\alpha d\beta d\gamma$ is unaltered by T .

Further, we have from (13),

$$\begin{aligned} \frac{1}{\sin \Theta} \frac{\partial}{\partial \Theta} \left(\sin \Theta \frac{\partial (F_M X)}{\partial \Theta} \right) + A_2 (F_M X) &= -l_2(l_2 + 1) F_M X + 2 \frac{\partial F_M}{\partial \Theta} \frac{dX}{d\Theta} \\ &+ F_M \frac{1}{\sin \Theta} \frac{d}{d\Theta} \left(\sin \Theta \frac{dX}{d\Theta} \right). \end{aligned} \quad (24)$$

Now we have seen that T converts F_M into $\pm \bar{F}_M$, so that, applying T and using (24), we have

$$\begin{aligned} \int F_M^* X \left[\frac{1}{\sin \Theta} \frac{\partial}{\partial \Theta} \left(\sin \Theta \frac{\partial (F_M X)}{\partial \Theta} \right) + A_1 (F_M X) \right] dv \\ = \int \bar{F}_M^* X \left[-l_1(l_1 + 1) \bar{F}_M X + 2 \frac{\partial \bar{F}_M}{\partial \Theta} \frac{dX}{d\Theta} + \bar{F}_M \frac{1}{\sin \Theta} \frac{d}{d\Theta} \left(\sin \Theta \frac{dX}{d\Theta} \right) \right] dv \\ = \int F_M^* X \left[-l_1(l_1 + 1) F_M X + 2 \frac{\partial F_M}{\partial \Theta} \frac{dX}{d\Theta} + F_M \frac{1}{\sin \Theta} \frac{d}{d\Theta} \left(\sin \Theta \frac{dX}{d\Theta} \right) \right] dv, \end{aligned} \quad (25)$$

the last step following by applying the transformation T^{-1} .

† I.e. in the original α, β, γ , (α, β) are polar co-ordinates of electron 1; in the new α, β, γ , (α, β) are polar co-ordinates of electron 2. The co-ordinate Θ is unaltered, but enters into the formulae of transformation.

Again, from (19) we have

$$\begin{aligned} \int |F_M|^2 dv &= (2\pi)^2 \sum_{M'=-L}^L c_{M'}^2 [P_l^{M'}(x)]^2 \int_0^\pi [y_{MM'}^L(\beta)]^2 \sin \beta d\beta \\ &= \frac{8\pi^2}{2L+1} \sum_{M'=-L}^L c_{M'}^2 [P_l^{M'}(x)]^2, \dagger \end{aligned} \quad (26)$$

and similarly,

$$\int F_M^* \frac{\partial F_M}{\partial \theta} dv = -\frac{8\pi^2}{2L+1} (1-x^2)^{\frac{1}{2}} \sum_{M'=-L}^L c_{M'}^2 P_l^{M'}(x) \frac{dP_l^{M'}(x)}{dx}. \quad (27)$$

Lastly, we can suppose the functions P_1, P_2, X normalized in any way we please without loss of generality. We shall suppose them normalized so that

$$\int_0^\infty P_1^2 dr = \int_0^\infty P_2^2 dr = \int |F_M|^2 X^2 dx dv = 1. \quad (28)$$

Using (24)–(28) in the variational principle (23), and taking into account the fact that the functions P_1, P_2, X in (20) are *arbitrary* functions, we find that these functions must satisfy the following differential equations:

$$\frac{d^2 P_1}{dr^2} + \left[\epsilon_1 + 2V(r) - \frac{l_1(l_1+1) - C}{r^2} - 2U_1(r) \right] P_1 = 0, \quad (29)$$

$$\frac{d^2 P_2}{dr^2} + \left[\epsilon_2 + 2V(r) - \frac{l_2(l_2+1) - C}{r^2} - 2U_2(r) \right] P_2 = 0, \quad (30)$$

$$\frac{d}{dx} \left[(1-x^2) \frac{dX}{dx} \right] + 2\lambda(x) \frac{dX}{dx} + [\epsilon_3 - 2U_3(x)] X = 0, \quad (31)$$

where $\epsilon_1, \epsilon_2, \epsilon_3$ are constants of the nature of eigenvalue parameters, and the remaining symbols have the following meanings:

$$\begin{aligned} C &= \sum_{M'=-L}^L c_{M'}^2 \int_{-1}^1 X P_l^{M'}(x) \\ &\quad \times \left[2(1-x^2) \frac{dX}{dx} - \frac{dP_l^{M'}(x)}{dx} + P_l^{M'}(x) \frac{d}{dx} \left((1-x^2) \frac{dX}{dx} \right) \right] dx / N, \end{aligned} \quad (32)$$

$$U_1(r_1) = \sum_{M'=-L}^L c_{M'}^2 \int_{-1}^1 dx \int_0^\infty dr_2 \cdot (1/r_{12}) [P_2(r_2) P_l^{M'}(x) X]^2 / N, \quad (33)$$

† The fact that $\int_0^\pi [y_{MM'}^L(\beta)]^2 \sin \beta d\beta = 2/(2L+1)$ may be seen most simply by utilizing one of the general orthogonality relations of irreducible unitary representations of continuous groups (Wigner 1931, p. 110, formula (11)), with the special form given by (7) for the representation.

$$U_2(r_2) = \sum_{M'=-L}^L c_{M'}^2 \int_{-1}^1 dx \int_0^\infty dr_1 \cdot (1/r_{12}) [P_1(r_1) P_l^{M'}(x) X]^2 / N, \quad (34)$$

$$N = \sum_{M'=-L}^L c_{M'}^2 \int_{-1}^1 [P_l^{M'}(x) X]^2 dx, \quad (35)$$

$$\lambda(x) = (1-x^2) \sum_{M'=-L}^L c_{M'}^2 P_l^{M'}(x) \frac{dP_l^{M'}(x)}{dx} \bigg/ \sum_{M'=-L}^L c_{M'}^2 [P_l^{M'}(x)]^2, \quad (36)$$

$$U_3(x) = \int_0^\infty \int_0^\infty (1/r_{12}) [P_1(r_1) P_2(r_2)]^2 dr_1 dr_2 \bigg/ \int_0^\infty (P_1^2 + P_2^2)/r^2 \cdot dr. \quad (37)$$

We have inserted the normalization factor for the function X from (28), which gives simply $8\pi^2 N/(2L+1) = 1$, in order that only the *ratios* of the c 's may appear in the final results. In (32) we can, of course, substitute for $(d/dx)[(1-x^2)dX/dx]$ from (31). The integrals containing $1/r_{12}$ may be further simplified by utilizing the well-known expansion

$$1/r_{12} = \sum_{n=0}^{\infty} (r_a^n/r_b^{n+1}) P_n(x), \quad (38)$$

where $r_a = \min(r_1, r_2)$, $r_b = \max(r_1, r_2)$.

Equations (29), (30), (31) are the fundamental equations of the present method (without exchange). From (32)–(37) we see that P_1, P_2, X are independent of M , as required. Further, (36) and (37) show that X is unaltered by interchanging the two electrons, this being true for the function $\lambda(x)$, since it is equal, except for a constant factor, to $\int F_M^* \frac{\partial F_M}{\partial \Theta} dv \bigg/ \int |F_M|^2 dv$, and both these integrals are unaltered by the interchange, as is seen by applying the transformation T .

Equations (29), (30) are similar to the radial wave equations occurring in the s.c.f. method without exchange, but they differ in the addition of the term C/r^2 (equivalent to an alteration of the azimuthal quantum number) and also in the fact that the potentials U_1, U_2 arising from the electronic interaction depend on the angular function X , and also depend on L as well as l_1, l_2 . The principal quantum number n would be introduced in the usual way, that eigenfunction of (29), for instance, being taken which has $(n_1 - l_1 - 1)$ zeros (excluding the origin and infinity).

Equation (31) must be solved subject to the condition that the function X remains bounded throughout the interval $-1 \leq x \leq 1$, it bears some resemblance to Legendre's equation. Which eigenfunction should be taken can be seen by considering the case of zero interaction, $1/r_{12} = 0$. For in this case, we know that the required solution is simply $X = \text{constant}$, i.e. that

eigenfunction which has no zeros in the interval $(-1, 1)$; from continuity considerations, we therefore assume that this is so also in the general case.

Equation (31) has regular singularities at the points $x = \pm 1$, but the function $\lambda(x)$ may introduce additional singularities, which would usually be inconvenient. Since, however, the functions $P_l^{M'}(x)$, for different values of M' , have no zeros in common other than $x = \pm 1$, an inspection of (36) and (17) shows that such extra singularities can be introduced only in the case of S terms or even P terms. These are two of the cases mentioned in the last section where the function F_M in (20) is equal to the product of a function of x and a function of α, β, γ , and, as remarked before, the "x" part of F can then be absorbed into the function X . Effecting this change in the cases referred to, we find that (31) is to be altered as follows:

$$\text{S terms: } \frac{d}{dx} \left[(1-x^2) \frac{dX}{dx} \right] + [\epsilon_3 - U_3(x)] X = 0, \quad (31')$$

$$\text{Even P terms: } \frac{d}{dx} \left[(1-x^2) \frac{dX}{dx} \right] + \left[\epsilon_3 - \frac{1}{1-x^2} - U_3(x) \right] X = 0. \quad (31'')$$

By the continuity argument used in discussing (31), we see that that eigenfunction of (31') must be taken which has l nodes in the interval $(-1, 1)$, and that eigenfunction of (31'') which has $(l-1)$ nodes. The normalization for X can now conveniently be taken to be

$$\int_{-1}^1 X^2 dx = 1, \quad (39)$$

and (32)–(34) are to be modified as follows:

$$C = \int_{-1}^1 [l(l+1) - \epsilon_3 + U_3(x)] X^2 dx, \quad (32')$$

$$U_1(r_1) = \int_{-1}^1 dx \int_0^\infty dr_2 \cdot (1/r_{12}) [P_2(r_2) X]^2, \quad (33')$$

$$U_2(r_2) = \int_{-1}^1 dx \int_0^\infty dr_1 \cdot (1/r_{12}) [P_1(r_1) X]^2. \quad (34')$$

These hold for either of the above cases with $l_1 = l_2 = l$, and ϵ_3 in (32') referring to the parameter occurring in (31') or (31''). If $l_2 = 1$, $L = l_1$, F_M is also equal to the product of a function of x and a function of α, β, γ ; but in this case no additional singularities are introduced into (31), and the above process is not necessary.† If, however, we do absorb the "x" part

† The same remark applies to the cases of S or even P terms, with $l_1 = l_2 = 1$. The case $l_2 = 0$, in which case F_M is independent of x , can be dealt with as a special case of (31), $\lambda(x)$ being then equal to zero.

of F_M into the function X , we use (31'') and (32')-(34'), with l put equal to 1. These exceptional cases are actually seen to cover some of the commonest configurations.

The fundamental equations of the present method can be solved by the usual process of successive approximations. Starting with the radial wave functions of the s.c.f., (31), or its alternatives (31') or (31''), can be solved by the usual numerical methods. With this solution for X , the radial equations can then be resolved, and so on. A few terms of (38) should, in practice, suffice for the calculation of integrals involving $1/r_{12}$, and the labour demanded over and above that required for the ordinary s.c.f. method should not be too excessive. Equations (31') and (31'') can be replaced by variational problems in the usual way, if a simple analytical approximation to X with adjustable constants is found to be feasible; (31) must first be made self-adjoint before being so replaced.

We can take account of polarization of the atom core by assuming a wave function which is a product of one-electron central-field wave functions of *all* the electrons, but with the generalization already considered as far as the *valence* electrons are concerned, i.e. assume

$$\psi = \psi_M^L(1, 2) \psi_3(3) \dots \psi_N(N), \quad (40)$$

where the core electrons are numbered 3, ..., N . $\psi_M^L(1, 2)$ is again given by (20), and ψ_3, \dots, ψ_N are of the usual form of one-electron central-field wave functions. If the variational principle be then applied, and the resulting equations averaged over the magnetic quantum numbers of the core electrons (as is necessary, since these quantum numbers have no physical significance; for the resultant state—see below), we obtain the previous equations (29), (30), with

$$V(r) = \frac{Z}{r} - \sum_{i=3}^N \int_0^\infty \frac{[P_i(r_i)]^2}{\max(r, r_i)} dr_i \quad (Z = \text{nuclear charge}),$$

while (31) is unaltered, and the equations for the radial functions of the core electrons are of exactly the same form as in the s.c.f., without exchange, but since the radial functions P_1, P_2 of the valence electrons will not be the same as in the s.c.f., the core wave functions will not be the same either.

A point may be mentioned here connected with the equations of the s.c.f. (without exchange) as ordinarily used. If a product of one-electron central-field wave functions of the usual type be assumed in the variational principle, the potential of the field, due to the other electrons, in which each electron moves, depends on the magnetic quantum numbers of the individual electrons, as pointed out by Fock (1930) and Slater (1930). But these

magnetic quantum numbers have no physical significance for the resultant state; if an *average* over magnetic quantum numbers is taken, we obtain just the Hartree potentials.† The correct spatial symmetry is not here allowed for, however; if it is allowed for (by using the generalization of (18) to any number of electrons), the field for each separate electron depends not only on the electronic configuration but also on the resultant state (i.e. the quantum number L), but any dependence on magnetic quantum numbers is eliminated. For the case of two electrons, the correct equations can be deduced from (29) and (30) by putting $X = \text{constant}$. The corrections to the Hartree equations thus obtained, however, are in general of the same order of magnitude as those arising from exchange, so that there is no point in taking them into account without at the same time taking into account the exchange terms (Slater 1930). It can, moreover, be shown that the *average* of the correct potentials (an " L " state being weighted with the factor $(2L+1)$) gives just the Hartree potential.

In the s.c.f. *with* exchange, as used by Hartree and Hartree (1935, 1936*a, b*), the proper spatial symmetry *is*, in effect, allowed for, and the equations are, in fact, different for the different states arising from the same electronic configuration. It may be pointed out, however, that the assumption of orthogonality made with regard to the individual wave functions is, in some cases, a restriction on the form of the wave function as a whole. But if this restriction be abandoned, the equations would be thereby complicated, without, probably, much increase in accuracy being obtained.

4—THE EQUATIONS WITH EXCHANGE

One further improvement in the present method, in line with what has already been done in the s.c.f. method, is possible: namely, to take account of exchange degeneracy in the assumed form of the wave function. In the case of a two-electron configuration, the proper form is

$$\Psi = \psi_M^L(1, 2) \pm \psi_M^L(2, 1), \quad (41)$$

where the $+$ sign is to be taken for singlets, the $-$ for triplets.‡

Using (41), (20), (22), (28) in the variational principle (23), and using

† Condon and Shortley (1935, p. 356). There would seem to be no point in attempting to take this dependence on magnetic quantum numbers into account in the equations, as has been done, e.g. by Kennard and Ramberg (1934). See also Stevenson (1937).

‡ In the case of *equivalent* electrons, we see from (22) that the singlet wave function vanishes if L is odd, and the triplet wave function if L is even—in agreement with Pauli's principle.

equations analogous to (24)–(27) to simplify the analysis, the following equations result:

$$\begin{aligned} \frac{d^2 P_1}{dr^2} + \left[\epsilon_1 + 2V(r) - \frac{l_1(l_1+1) - C}{r^2} - 2U_1(r) \right] P_1 + K \frac{d^2 P_2}{dr^2} \\ + \left[\epsilon_{12} + 2KV(r) - \frac{(K/2)(l_1(l_1+1) + l_2(l_2+1)) - C'}{r^2} - 2U_{12}(r) \right] P_2 = 0, \\ \frac{d^2 P_2}{dr^2} + \left[\epsilon_2 + 2V(r) - \frac{l_2(l_2+1) - C}{r^2} - 2U_2(r) \right] P_2 + K \frac{d^2 P_1}{dr^2} \\ + \left[\epsilon_{12} + 2KV(r) - \frac{(K/2)(l_1(l_1+1) + l_2(l_2+1)) - C'}{r^2} - 2U_{12}(r) \right] P_1 = 0, \\ \frac{d}{dx} \left[(1-x^2) \frac{dX}{dx} \right] [1 + Af(x)] + 2[\lambda(x) + A\lambda^{(1)}(x)] \frac{dX}{dx} \\ + [\epsilon_3 + \epsilon'_3 f(x) - 2U_3(x) - 2U_3^{(1)}(x)] X = 0, \end{aligned}$$

where ϵ_1 , ϵ_2 , ϵ_3 are eigenvalue parameters,† C , $U_1(r)$, $U_2(r)$, $\lambda(x)$, $U_3(x)$ have the same meanings as before, and the remaining symbols are defined as follows:

$$\begin{aligned} K &= \alpha a \sum_{M'=-L}^L c_{M'} c_{M'} \int_{-1}^1 P_{l_1}^{M'}(x) P_{l_1}^{M'}(x) X^2 dx / N, \\ a &= \int_0^\infty P_1 P_2 dr, \quad \alpha = \pm (-1)^{L+l_1+l_2} \quad (+ \text{ for singlets, } - \text{ for triplets}), \\ A &= 2\alpha a \int_0^\infty P_1 P_2 / r^2 \cdot dr / \int_0^\infty (P_1^2 + P_2^2) / r^2 \cdot dr, \\ C' &= \alpha a \sum_{M'=-L}^L c_{M'} \bar{c}_{M'} \int_{-1}^1 P_{l_1}^{M'}(x) X \\ &\quad \times \left[2(1-x^2) \frac{dX}{dx} \frac{dP_{l_1}^{M'}(x)}{dx} + P_{l_1}^{M'}(x) \frac{d}{dx} \left((1-x^2) \frac{dX}{dx} \right) \right] dx / N, \\ (a/K) \epsilon_{12} &= \int_0^\infty \left\{ (\epsilon_3/r^2) (P_1^2 + P_2^2) + 2V(P_1 P_2 - P_1^2 - P_2^2) - P_1' P_2' + P_1'^2 + P_2'^2 \right. \\ &\quad \left. + [l_1(l_1+1) P_1^2 + l_2(l_2+1) P_2^2 - (C'/K) P_1 P_2] / r^2 \right\} dr, \end{aligned}$$

† They are not necessarily the same as the parameters occurring in the equations without exchange. That there are three independent eigenvalue parameters can be seen by considering the variational problem in the form $\delta \int \Psi^* H \Psi d\tau = 0$, subject to three normalization conditions for the three functions P_1 , P_2 , X . The normalization for X would then be different from that in (27), but the equations are given in a form independent of this normalization. P_1 , P_2 are supposed normalized to unity as before.

$$\begin{aligned}
& (\epsilon'_2 - a^2 \epsilon_2) \int_0^\infty (P_1^2 + P_2^2)/r^2 \cdot dr \\
& = a \int_0^\infty [4VP_1P_2 - (l_1(l_1+1) + l_2(l_2+1))/r^2 \cdot P_1P_2 - 2P'_1P'_2] dr \\
& \quad + a^2 \int_0^\infty [P_1'^2 + P_2'^2 - (l_1(l_1+1)P_1^2 + l_2(l_2+1)P_2^2)/r^2 - 2V(P_1^2 + P_2^2)] dr, \\
U_{12}(r_1) &= \alpha \sum_{M'=-L}^L c_{M'} \bar{c}_{M'} \int_{-1}^1 dx \int_0^\infty dr_2 \\
& \quad \times (1/r_{12}) P_1(r_1) P_2(r_2) P_{l_1}^{M'}(x) P_{l_2}^{M'}(x) X^2/N, \\
f(x) &= \mu(x)/\chi(x), \quad \mu(x) = \sum_{M'=-L}^L c_{M'} \bar{c}_{M'} P_{l_1}^{M'}(x) P_{l_2}^{M'}(x), \\
\chi(x) &= \left[\sum_{M'=-L}^L c_{M'}^2 (P_{l_1}^{M'}(x))^2 \cdot \sum_{M'=-L}^L \bar{c}_{M'}^2 (P_{l_2}^{M'}(x))^2 \right]^{\frac{1}{2}}, \\
\lambda^{(1)}(x) &= (1-x^2) \sum_{M'=-L}^L c_{M'} \bar{c}_{M'} P_{l_1}^{M'}(x) \frac{dP_{l_2}^{M'}(x)}{dx} / \chi(x), \\
U_3^{(1)}(x) &= \alpha f(x) \iint_0^\infty (1/r_{12}) P_1(r_1) P_2(r_1) P_1(r_2) P_2(r_2) dr_1 dr_2 / \int_0^\infty (P_1^2 + P_2^2)/r^2 \cdot dr,
\end{aligned}$$

while for N , we now use, instead of (35),†

$$N = \left[\sum_{M'=-L}^L c_{M'}^2 (P_{l_1}^{M'}(x))^2 \cdot \sum_{M'=-L}^L \bar{c}_{M'}^2 (P_{l_2}^{M'}(x))^2 \right]^{\frac{1}{2}}.$$

The new differential equations are considerably more complicated than those without exchange, though the new radial equations bear some resemblance to those of the s.c.f. with exchange. A numerical solution, though feasible, would be laborious.† For numerical work, it would probably be advantageous first to solve the left-hand members of the radial equations with respect to the second-order derivatives. The modifications necessary when the "x" part of F_M can be absorbed into the function X , analogous to those already considered in § 3, can be obtained without difficulty.

† In order that only the ratios of the c 's or \bar{c} 's may occur in the final results. The two expressions for N are equal, since $\int |F_M|^2 dv = \int |\bar{F}_M|^2 dv$ by the transformation T . We have made a similar alteration in the expression for $\chi(x)$.

‡ If $l_1 = l_2$, the equations simplify considerably, for then $f(x) = 1$, $\lambda^{(1)}(x) = \lambda(x)$, $K = aa$, $C' = aaC$. In this case, it would also be permissible to impose the condition that P_1 and P_2 are orthogonal ($a = 0$), which would further simplify the equations; ϵ_{11} would then have to be considered as an *independent* eigenvalue parameter.

The above does not, of course, take account of exchange involving the core electrons. This can be done as follows: writing $\Psi_K^L(1, 2)$ for the wave function (41) when the appropriate spin factors are included, the required wave function for the whole atom can be written as a linear combination of terms of the form

$$\Psi_K^L(\alpha_1, \alpha_2) \Delta(\alpha_3, \dots, \alpha_N),$$

where $\alpha_1, \dots, \alpha_N$ is some permutation of $1, \dots, N$, and Δ is the Slater determinant for the core electrons. The wave function thus contains the angular function X with the arguments $x_{ij} = \cos(\mathbf{r}_i, \mathbf{r}_j)$, ($i, j = 1, \dots, N$; $i \neq j$). The variational principle then yields equations which are a generalization of those considered at the beginning of this section.

5—THE TOTAL ENERGY

In considering the expression for the total energy (measured from double ionization), we shall neglect polarization effects. The energy is given by

$$E = \int \Psi^* H \Psi d\tau / \int |\Psi|^2 d\tau.$$

With the expression (41) for Ψ , and utilizing the equations (29) and (30) *without* exchange, we find†

$$E = \epsilon_1 + \epsilon_2 + \frac{2}{1 + Ka} \\ \times \int_0^\infty \left[U_{12} P_1 P_2 - U_1 P_1^2 - K(U_1 + U_2) P_1 P_2 + \frac{KC - C'}{r^2} P_1 P_2 \right] dr,$$

where K, a, C' are as defined in the last section; but ϵ_1, ϵ_2 and the functions P_1, P_2, X refer, of course, to the equations *without* exchange, namely (29)–(31).

For the corresponding equations *with* exchange, we find the simpler expression

$$E = 1/(1 + Ka) \cdot \left[\epsilon_1 + \epsilon_2 + 2a\epsilon_{12} - 2 \int_0^\infty (U_1 P_1^2 + U_{12} P_1 P_2) dr \right],$$

the parameters and functions now referring to the equations *with* exchange.

† In the case of *equivalent* electrons, we find simply

$$E = \epsilon_1 + \epsilon_2 - 2 \int_0^\infty U_1 P_1^2 dr$$

for the states allowed by Pauli's principle.

I am indebted to Professor Hartree, who has kindly given me helpful criticism on several points in this paper.

6—SUMMARY

The form of the wave function assumed in the ordinary method of the self-consistent field is generalized so as to allow for the correct spatial symmetry corresponding to a given term arising from a given configuration of the two valence electrons, and, in addition, an adjustable function of the angle between the two radii vectores is inserted as a multiplying factor. Application of the variational method then yields differential equations for the radial functions which are similar to those of the ordinary theory, together with an equation for the angular function.

The equations are also given for the case when exchange degeneracy is taken into account. The equations with exchange prove to be quite complicated, but those without exchange should be soluble without too much labour when the corresponding solutions of the ordinary self-consistent field equations are available.

The application of the method when the core electrons are taken into account in the assumed form of the wave function is also briefly considered.

REFERENCES

- Breit, G. 1930 *Phys. Rev.* **35**, 569-78.
Condon, E. U. and Shortley, G. H. 1935 "The theory of atomic spectra." Camb. Univ. Press.
Fock, V. 1930 *Z. Phys.* **61**, 126-48.
Hartree, D. R. and Hartree, W. 1935 *Proc. Roy. Soc. A*, **150**, 9-33.
— — 1936a *Proc. Roy. Soc. A*, **154**, 588-607.
— — 1936b *Proc. Roy. Soc. A*, **156**, 45-62.
Hylleraas, E. 1928 *Z. Phys.* **48**, 469-94.
— 1929 *Z. Phys.* **54**, 347-66.
Kennard, E. H. and Ramberg, E. 1934 *Phys. Rev.* **46**, 1034-40.
Slater, J. C. 1930 *Phys. Rev.* **35**, 210-11.
Stevenson, A. F. 1937 *Phys. Rev.* **51**, 285-7.
Weyl, H. 1931 "Theory of Groups and Quantum Mechanics." Methuen and Co.
Wigner, E. 1931 "Gruppentheorie." Braunschweig: F. Viewig und Sohn
-

INDEX TO VOLUME CLX (A)

- Adsorption of gases, and the equation of the liquid state (Baly), 465.
 Adsorption on measured surfaces of vitreous silica (Palmer), 254.
 Alpha particles, scattering of, in helium, hydrogen and deuterium (Mohr and Pringle), 190.
 Antiplane stress in an elastic solid (Filon), 137.
 Atkins (W. R. G.), Ball (N. G.) and Poole (H. H.) The photo-electric measurement of the diurnal variations in daylight in temperate and tropical regions, 526.
 Atomic constants, fundamental, values of (von Friesen), 424.
- Ball (N. G.) *See* Atkins, Ball and Poole
 Baly (E. C. C.) The adsorption of gases and the equation of the liquid state, 465.
 Bartlett (M. S.) Properties of sufficiency and statistical tests, 268.
 Beilby layer, formation of (Bowden and Hughes), 375.
 Beta-rays, nuclear, of radium D (Richardson and Leigh-Smith), 454.
 Bijl (A.) *See* Michels, Bijl and Michels.
 Blackett (P. M. S.) and Wilson (J. G.) The energy loss of cosmic ray particles in metal plates, 304.
 Blaisse (B.) *See* Michels, Blaisse and Michels.
 Bolland (J. L.) *See* Melville and Bolland.
 Bolland (J. L.) *See* Melville, Bolland and Roxburgh.
 Bowden (F. P.) and Hughes (T. P.) Physical properties of surfaces. IV—Polishing, surface flow, and the formation of the Beilby layer, 575
 Buckingham (R. A.) The quantum theory of atomic polarization. I—Polarization by a uniform field. II—The van der Waals energy of two atoms, 94, 113.
- Cadmium, magneto-resistance effect in, at low temperatures (Milner), 207.
 CO₂ between 0 and 150° and up to 3000 atm., series evaluation of the isotherm data of (Michels and Michels), 348.
 CO₂, isotherms of (Michels, Blaisse and Michels), 358.
 CO₂, thermodynamic properties of (Michels, Bijl and Michels), 376
 Cosmic-ray particles, energy loss of, in metal plates (Blackett and Wilson), 304.
 Crystals, ionic, theory of electrical breakdown in (Frohlich), 230.
- Daunt (J. G.) and Mendelssohn (K.) Equilibrium curve and entropy difference between the supraconductive and the normal state in Pb, Hg, Sn, Ta, and Nb, 127.
 Daylight, diurnal variations in, in temperate and tropical regions (Atkins, Ball and Poole), 526.
 Deuterium and phosphine, mercury photosensitized exchange reaction of (Melville and Bolland), 384.
 Dirac (P. A. M.) Complex variables in quantum mechanics, 48.
- Electrical breakdown, theory of, in ionic crystals (Frohlich), 230.

- Filon (L. N. G.) On antiplane stress in an elastic solid, 137.
- Fowler (R. H.) and Smithells (C. J. S.) A theoretical formula for the solubility of hydrogen in metals, 37.
- Fraser (R. G. J.) and Jewitt (T. N.) The ionization potentials of the free radicals methyl and ethyl, 563.
- Freezing-points of alloys, the accurate determination of (Hume-Rothery and Reynolds), 282.
- Friesen (Sten von) On the values of fundamental atomic constants, 424.
- Fröhlich (H.) Theory of electrical breakdown in ionic crystals, 230.
- Gas-air mixtures, inflammable, influence of pressure on the spontaneous ignition of (Kane and Townend), 174.
- Gases, adsorption of, and the equation of the liquid state (Baly), 465.
- Gravitation, inverse square law of, II, III (Milne), 1, 24.
- H₂, band systems (Richardson), 487.
- Howell (H. G.) The spectrum of thallium fluoride, 242.
- Howell (O. R.) and Warne (H.) The transport numbers of the ions in solutions of silver dodecyl sulphate, 440.
- Hughes (T. P.) See Bowden and Hughes.
- Hume-Rothery (W.) and Reynolds (P. W.) The accurate determination of the freezing-points of alloys, and a study of valency effects in certain alloys of silver, 282.
- Hydrogen, solubility of, in metals, theoretical formula (Fowler and Smithells), 37.
- Ionization potentials of the free radicals methyl and ethyl (Fraser and Jewitt), 563.
- Ions, transport numbers of, in solutions of silver dodecyl sulphate (Howell and Warne), 440.
- Jeffreys (H.) On the relation between direct and inverse methods in statistics, 325.
- Jewitt (T. N.) See Fraser and Jewitt.
- Kane (G. P.) and Townend (D. T. A.) The influence of pressure on the spontaneous ignition of inflammable gas-air mixtures—The simpler olefines, 174.
- Leigh-Smith (A.) See Richardson and Leigh-Smith.
- Linnett (J. W.) See Thompson and Linnett.
- Melville (H. W.), Bolland (J. L.) and Roxburgh (H. L.) The photochemical decomposition and oxidation of trideutero-phosphine, 406.
- Melville (H. W.) and Bolland (J. L.) The mercury photosensitized exchange reaction of deuterium and phosphine, 384.
- Mendelssohn (K.) See Daunt and Mendelssohn.
- Michels (A.) and Michels (C.) Series evaluation of the isotherm data of CO₂ between 0 and 150° C. and up to 3000 atm., 348.
- Michels (A.), Bijl (A.) and Michels (C.) Thermodynamic properties of CO₂ up to 3000 atm. between 25 and 150° C., 376.

- Michels (A.), Blaisse (B.) and Michels (C.) The isotherms of CO_2 in the neighbourhood of the critical point and round the coexistence line, 358.
- Michels (C.) *See* Michels and Michels.
- Michels (C.) *See* Michels, Bijl and Michels.
- Michels (C.) *See* Michels, Blaisse and Michels.
- Milne (E. A.) The inverse square law of gravitation—II, III, 1, 24.
- Milner (C. J.) The magneto-resistance effect in cadmium at low temperatures, 207.
- Mohr (C. B. O.) and Pringle (G. E.) The scattering of alpha-particles in helium, hydrogen and deuterium, 190.
- Olefines, the simpler, influence of pressure on spontaneous ignition (Kane and Townend), 174.
- Optical contact, further studies on (Rayleigh), 507.
- Palmer (W. G.) Adsorption on measured surfaces of vitreous silica—II, 254.
- Photochemical decomposition and oxidation of trideutero-phosphine (Melville and Bolland), 406.
- Photo-electric measurement of the diurnal variations in daylight (Atkins, Ball and Poole), 526.
- Photosensitized exchange reaction of deuterium and phosphine (Melville and Bolland), 384.
- Polarization, quantum theory of atomic (Buckingham), 94, 113.
- Polishing, surface flow and the formation of the Beilby layer (Bowden and Hughes), 575.
- Polyatomic molecules, photochemistry of (Thompson and Linnett), 539.
- Poole (H. H.) *See* Atkins, Ball and Poole.
- Pringle (G. E.) *See* Mohr and Pringle.
- Proteins, pattern of (Wrinch), 59.
- Quantum mechanics, complex variables in (Dirac), 48.
- Quantum theory of atomic polarization (Buckingham), 94, 113.
- Radium D, nuclear β -rays of (Richardson and Leigh-Smith), 454.
- Rayleigh, Lord The surface layer of polished silica and glass with further studies on optical contact, 507.
- Relativity, general, gravitational constant in (Synge), 187.
- Reynolds (P. W.) *See* Hume-Rothery and Reynolds.
- Richardson (H. O. W.) and Leigh-Smith (A.) The nuclear β -rays of radium D, 454.
- Richardson (O. W.) The band systems ending on the $1s\sigma 2s\sigma {}^1\Sigma_g ({}^1X_g)$ state of H_2 —Part I, 487.
- Roxburgh (H. L.) *See* Melville, Bolland and Roxburgh.
- Saha (M. N.) On the action of ultra-violet sunlight upon the upper atmosphere, 155.
- Schofield (R. K.) and Scott Blair (G. W.) The relationship between viscosity, elasticity and plastic strength of a soft material as illustrated by some mechanical properties of flour dough. IV—The separate contributions of gluten and starch, 87.

- Scott Blair (G. W.) *See* Schofield and Scott Blair.
- Self-consistent field for two-electron configurations, a generalization of the equations of the (Stevenson), 588.
- Silica and glass, polished, surface layer of (Rayleigh), 507.
- Silica, vitreous, adsorption on measured surfaces of (Palmer), 254.
- Silver dodecyl sulphate, transport numbers of the ions in solutions of (Howell and Warne), 440.
- Smithells (C. J. S.) *See* Fowler and Smithells.
- Spectra of absorption of polyatomic molecules containing alkyl radicals (Thompson and Linnett), 539.
- Spectrum of thallium fluoride (Howell), 242.
- Statistics, relation between direct and inverse methods in (Jeffreys), 325.
- Stevenson (A. F.) A generalization of the equations of the self-consistent field for two-electron configurations, 588.
- Sufficiency and statistical tests, properties of (Bartlett), 268.
- Supraconductive and normal state, equilibrium curve and entropy difference between (Daunt and Mendelssohn), 127.
- Surfaces, physical properties of (Bowden and Hughes), 575.
- Synge (J. L.) A criticism of the method of expansion in powers of the gravitational constant in general relativity, 187.
- Thallium fluoride, spectrum of (Howell), 242.
- Thompson (H. W.) and Linnett (J. W.) The absorption spectra and photochemistry of the polyatomic molecules containing alkyl radicals. Part V—Vibration frequencies and structure, 539.
- Townend (D. T. A.) *See* Kane and Townend.
- Trideuterophosphine, photochemical decomposition and oxidation of (Melville, Bolland and Roxburgh), 406.
- Upper atmosphere, action of ultra-violet sunlight upon (Saha), 155.
- Valency effects in certain silver alloys (Hume-Rothery), 282.
- Viscosity, elasticity and plastic strength of a soft material as illustrated by flour dough, IV (Schofield and Scott Blair), 87.
- Warne (H.) *See* Howell and Warne.
- Wilson (J. G.) *See* Blackett and Wilson.
- Wrinch (D. M.) On the pattern of proteins, 59.

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY,
NEW DELHI.

[illegible]